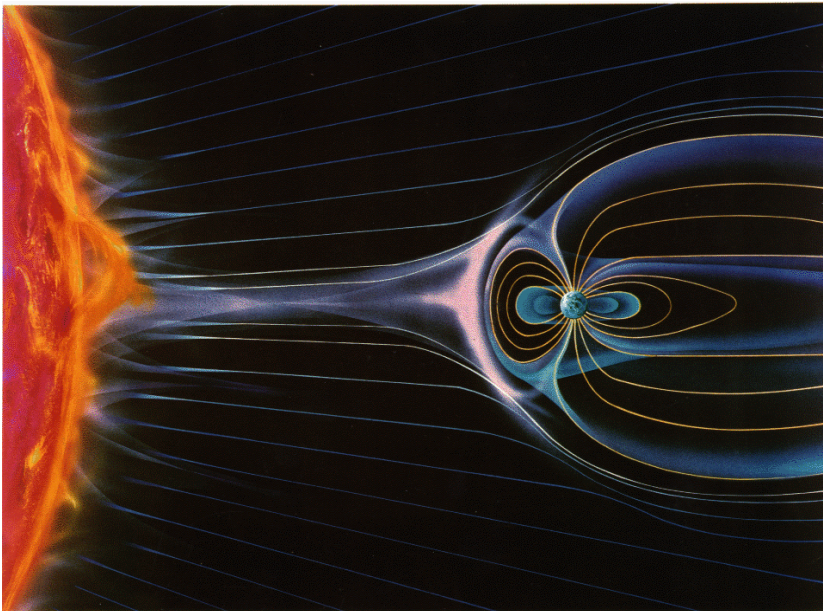

Introduction to the Space Environment

PH 2514



By R. C. Olsen - January 2003

CHAPTER 1	REVIEW OF SOME BASIC PHYSICS CONCEPTS	1
A	The Electromagnetic Spectrum	1
B	Sources of Electromagnetic Radiation	3
1	Line spectra	3
2	Black body radiation	11
3	X-rays - Brehmstrahlung	14
C	Kinetic Theory	15
1	Temperature and Kinetic Energy	15
2	Maxwellian Distributions	15
D	Plasmas	17
1	Phases of matter	17
2	Plasma Parameters	17
3	Definition of Plasma	20
4	Plasma Theories and Approximations	21
5	Orbit theory - Motion of Charged Particles	22
6	A few results from the magneto-hydrodynamic approximation	28
E	References	31
F	Problems	32
CHAPTER 2	THE SUN	35
A	Introduction	35
B	A few facts about the quiet sun	35
C	The Structure of the Sun	37
D	Solar Atmosphere	41
1	Chromosphere	42
2	Transition Region	43
3	The Corona	44
E	Solar Magnetism	49
1	Introduction	49
2	Sunspots	49
3	Sunspot Number and the Solar Cycle	51
F	Solar Activity	55
1	Prominences	55
2	Solar Flares	57
3	Coronal Mass Ejections	60
G	Energy Production in the Sun	61
H	References	63
I	Problems	64

CHAPTER 3	THE SOLAR WIND	67
A	Basic Characteristics	67
B	The "Supersonic" Solar Wind	69
C	The Interplanetary Magnetic Field (IMF)	72
D	Sector Structure	76
E	The Source of the Fast Solar Wind	78
F	Interplanetary Shocks	79
G	References	81
H	Problems	82
CHAPTER 4	THE GEOMAGNETIC FIELD	85
A	Introduction	85
1	Constituents of the Geomagnetic Field:	85
B	Main Geomagnetic Field	87
1	The Dipole Field	87
2	International Geomagnetic Reference Field	89
3	Magnetic Coordinate Systems	92
4	Dynamo Theory of the Main Field	94
C	Time Variations of the Geomagnetic Field	95
1	Diurnal Variations	95
2	Magnetic Storms	96
3	Magnetic Indices	99
D	References	102
E	Problems	102
CHAPTER 5	THE MAGNETOSPHERE	103
A	Introduction	103
B	Plasma Physics	104
1	The Magnetic Moment and Mirroring	104
2	Drifts	109
3	Characteristic Time Scales	110
4	Mcllwain L and Invariant Latitude	112
C	Magnetosphere structure	114
D	Bow Shock:	114
E	Magnetopause	115

F	Geosynchronous Orbit - The Plasmasphere and the Plasma Sheet	117
G	Magnetospheric Convection	121
H	The Radiation Belts	123
1	The Van Allen Radiation Belts	123
2	The Ring Current	126
3	The geocorona	127
4	Ring Current Decay and Composition	127
I	Problems	131
CHAPTER 6	THE ATMOSPHERE	135
A	Introduction	135
B	Law of Atmospheres: Pressure and Density	135
C	Temperature of the Atmosphere	139
D	Composition of the Atmosphere	141
E	Transmission of Electromagnetic Waves Through the Atmosphere	144
F	Problems	147
CHAPTER 7	THE IONOSPHERE	149
A	Introduction	149
B	Formation of an Ionospheric Layer (Chapman Theory)	152
1	Solar Intensity	153
2	Peak Altitude	153
3	Peak Production Rate	154
4	Production Rate	155
5		155
6	Loss Processes	156
C	Ionosphere morphology - Altitude Structure	156
1	D-region: 60 km to 90 km.	156
2	E layer: 90km-140 km:	157
3	F-region: 140km-1000km	159
D	Ionosphere morphology - Latitude Structure	162
E	Radiowave propagation in the Ionosphere	164
F	References	169
G	Problems	170

CHAPTER 8 ENVIRONMENTAL EFFECTS ON SPACE SYSTEMS 173

A	Spacecraft Charging	173
1	Introduction	173
2	Charging Mechanisms	173
3	Example of Spacecraft Charging – Geosynchronous Orbit	177
4	Design Considerations	179
B	Orbital Debris	180
1	Space Operations and Orbital Debris	180
2	Hypervelocity Impacts	188
3	Impact Probabilities.	190
4	Design Considerations	192
C	Surface Effects	193
1	Atomic Oxygen Effects	193
2	UV Degradation	195
3	Sputtering	195
4	Molecular Contamination	197
D	Homework	198

CHAPTER 9 ENVIRONMENTAL EFFECTS ON SPACE SYSTEMS - PART 2 RADIATION EFFECTS 199

A	Passage of ionization through matter.	199
1	Penetration of heavy charged particles	199
2	Interaction of Fast Electrons with Matter	202
3	Interaction of energetic photons with matter	204
4	Neutrons	205
B	Radiation Units	206
C	Effects of Space Radiation on Systems	208
1	Radiation Thresholds	209
2	Radiation Dose in the Belts	209
3	Solar Proton Events	220
4	Galactic Cosmic Rays	222
D	High-Altitude Nuclear Burst Effects	223
1	Characteristics Of A Nuclear Detonation	223
2	Ionization Effects In The Atmosphere	225
3	Electromagnetic Pulse (EMP)	226
4	Deposition Region Generated EMP	226
5	System Generated EMP/Internal EMP	227
6	Artificial Radiation Belts	228
E	Design Considerations Relative to Radiation	229

APPENDIX 1 USEFUL EQUATIONS 231

Chapter 1 Review of Some Basic Physics Concepts

A The Electromagnetic Spectrum

An understanding of the electromagnetic spectrum, and electromagnetic radiation is the first essential item we must address. Propagating electromagnetic energy may either be considered to be a special configuration of electric (\vec{E}) and magnetic (\vec{B}) fields traveling outward from a source or it may be described as a stream of particle-like objects called photons. In either case the radiation originates from electric charges undergoing accelerations or sudden transitions in their energy levels. The description of electromagnetic radiation (and its interaction with matter) as a wave phenomenon depends on the concepts of wavelength and frequency. These two quantities are related by

$$\lambda f = c \quad (\text{Eqn. 1.1})$$

Where λ = wavelength (in meters)
 f = Frequency in Hz (cycles/sec)
 c = Phase velocity of the wave (in m/sec)

For vacuum the value of $c = 2.998 \times 10^8$ (m/sec), an important constant of physics. The range of wavelengths which will be of interest to us covers a very large range of about 20 orders of magnitude as can be seen in Figure 1.1. (Initial study question: Where does x-band radar fall in this diagram?)

The connection between these two descriptions (wave-like and particle-like) for electromagnetic radiation generally merge in the concept of the photon, which is then often described as a particle with wave-like properties. The energy of an individual photon is given by

$$E = hf \text{ (Joules or eV)} \quad (\text{Eqn. 1.2})$$

where f = frequency of the EM wave (in Hz) and

$$h = \text{Planck's Constant} = \begin{cases} 6.626 \times 10^{-34} \text{ Joule - seconds} \\ 4.136 \times 10^{-15} \text{ eV - seconds} \end{cases}$$

The photon energy, E , is determined by the frequency of the electromagnetic radiation. The higher the frequency, the higher the energy. Photons move at the speed of light, as expected for electromagnetic radiation. They have zero rest mass, however, so the rules of special relativity are not violated.

The electron-volt (eV) is a convenient unit of energy and is related to the usual unit (Joule) by:

$$1 \text{ eV} = 1.602 \times 10^{-19} \text{ Joules} \quad (\text{Eqn. 1.3})$$

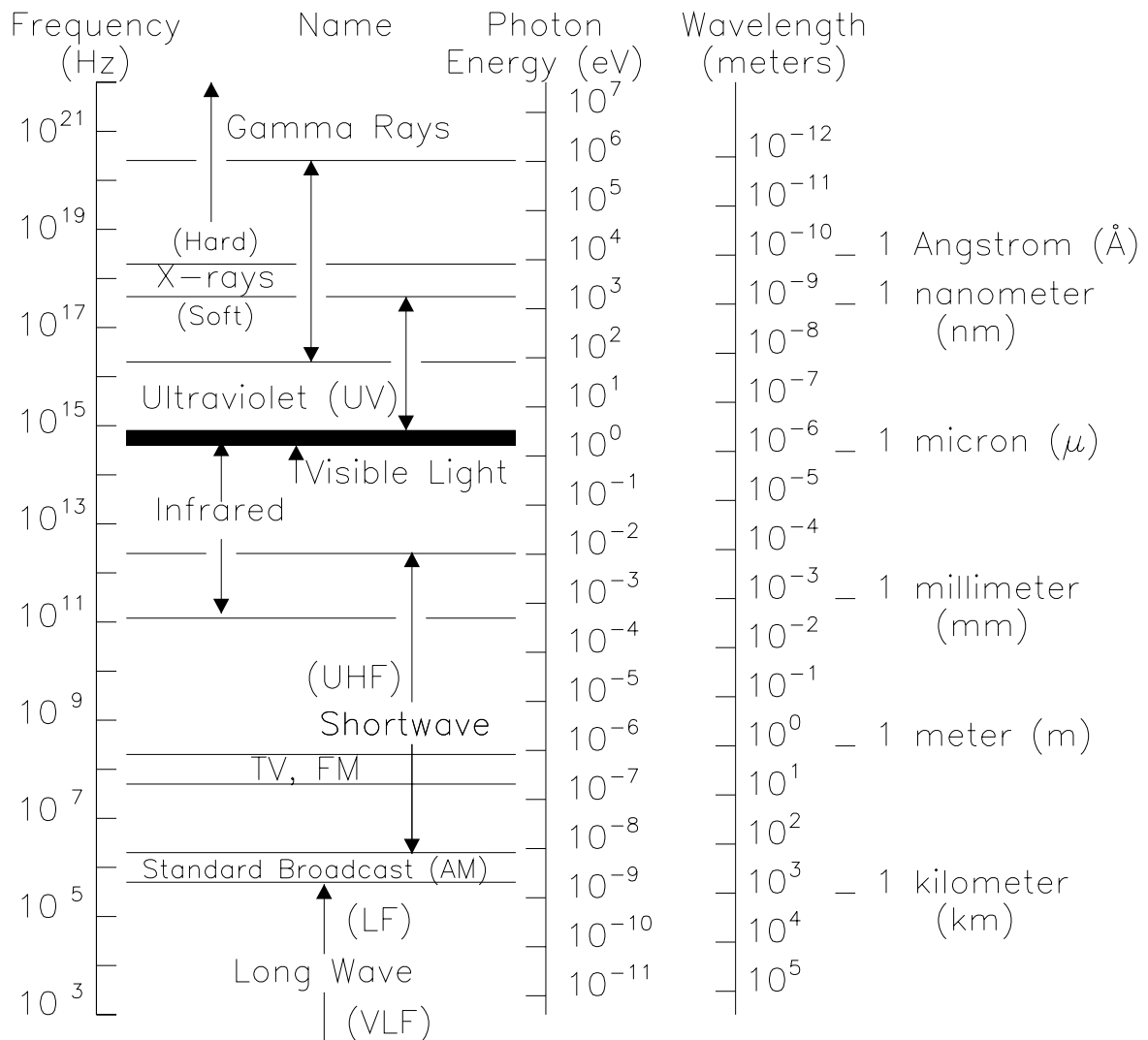


Figure 1.1 The spectrum of electromagnetic radiation

The question is often asked: What is electromagnetic radiation really? Is it a wave or is it a stream of particles? To answer this question we must perform some carefully designed experiment which will tell us. Depending on the experiment we find that we can get either answer, but only one at a time. It will also be found that in general the wave aspects dominate at frequencies below about 10^{15} Hz and the particle aspects at higher frequencies. In the visible part of the spectrum both descriptions are useful.

Example:

The energy of the photons making up an electromagnetic wave (light wave) in the visible part of the spectrum (green color) is

$$E = hf = (4.14 \times 10^{-15}) (6 \times 10^{14}) = 2.48 \text{ eV}$$

which is on the order of (or slightly less than) typical atomic binding energies.

Energies of typical x-ray photons are in the 10^4 to 10^5 eV range, while the photons of a 100 MHz radio signal are only about 4×10^{-7} eV.

B Sources of Electromagnetic Radiation

There are 3 major sources of electromagnetic radiation which are of interest to us in this course:

- (a) Individual atoms or molecules which radiate line spectra
- (b) Hot, dense bodies which radiate a continuous "black-body" spectrum
- (c) Energetic electrons which radiate x-rays when accelerated (or decelerated)

Each share a common element, in the end - that electromagnetic radiation is ultimately produced by electrons which are being accelerated (classically), or changing energy (quantum mechanics). The former case is most obvious in case c, the latter in case a. These sources of radiation are now considered in somewhat more detail.

1 Line spectra

Line spectra are emitted by single atoms or molecules. An atom or molecule which is reasonably isolated (such as in a gas at ordinary temperatures and pressures) will **radiate** a discrete set of frequencies called a line spectrum, as shown in Figure 1.2. If, on the other hand, we pass radiation having a continuous spectrum of frequencies through a gas we find that a discrete set of frequencies is **absorbed** by the gas leading to a spectrum of discrete absorption lines

The wavelengths radiated (absorbed) are characteristic of the particular atom or molecule and thus represent a powerful tool for determining the composition of radiating (or absorbing) gases. Much of our knowledge of the chemical composition of stars (including our sun) comes from detailed analysis of such line spectra.

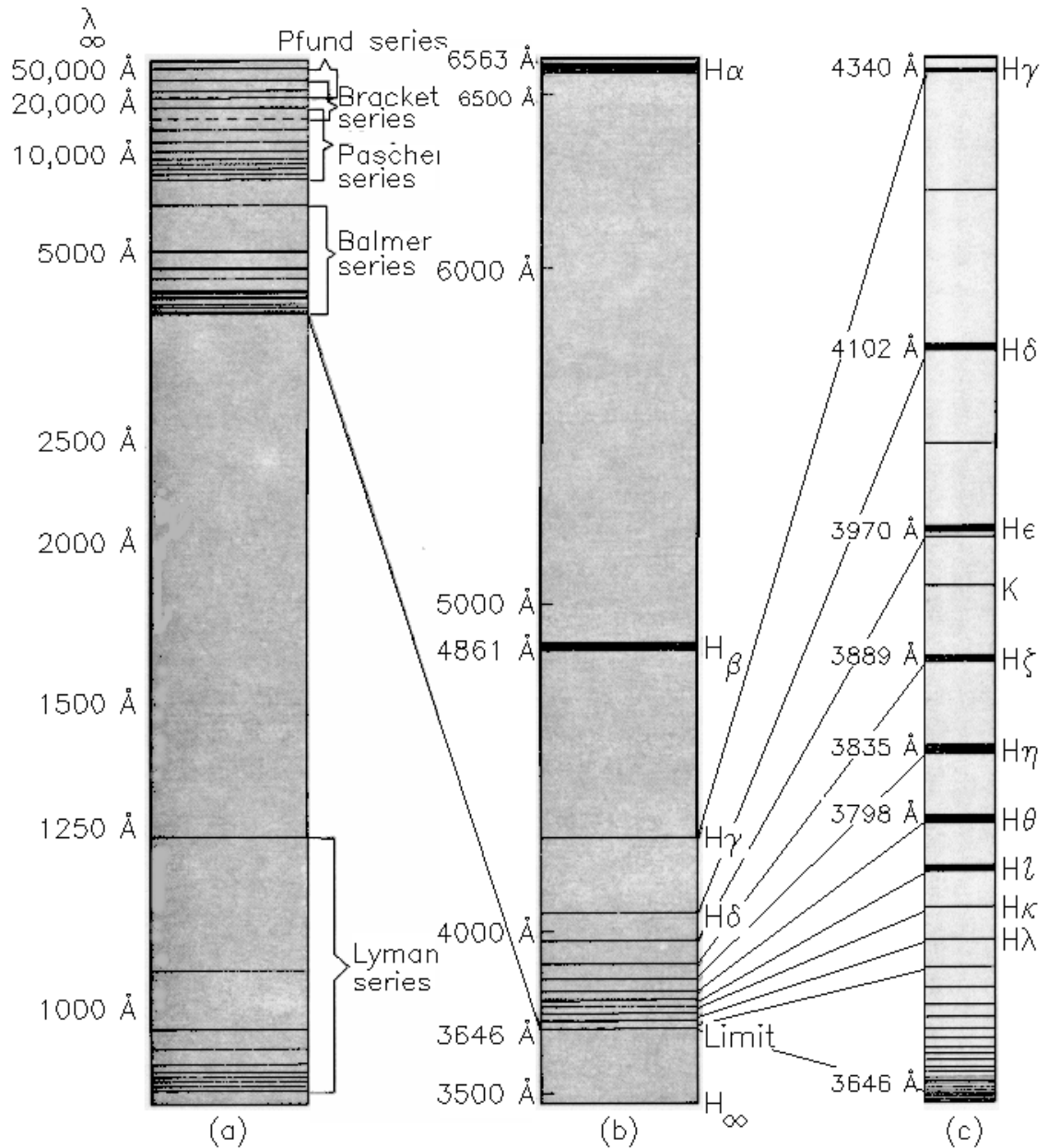


Figure 1.2 (a) The spectrum of atomic hydrogen consists of the Lyman series in the ultraviolet, the Balmer series in the visible, and several series in the infrared region. (b) The Balmer series in greater detail. (c) A portion of the spectrum of the star ζ Tauri showing more than 20 lines of the Balmer series.

Original citation: A. W. Smith and J. N. Cooper, "Elements of Physics", 7th edition, copyright 1964, McGraw Hill Book Co.

Found in: Introduction to Modern Physics, F. K. Richtmyer, E. H. Kennard, and John N. Cooper, page 229, 6th edition, 1969.

Derivation of the Bohr Atom

The existence of line spectra can be explained by means of the first 'quantum' model of the atom, developed by Bohr in 1913. Although the Bohr model of the hydrogen atom was eventually replaced, it yields the correct values for the observed spectral lines, and gives a substantial insight into the structure of atoms in general. The following derivation has the objective of obtaining the energy levels of the Bohr atom. If we can obtain the energy levels, we can reproduce the hydrogen atom spectra.

The derivation proceeds with three major elements: first, use force balance to relate the velocity to the radius of the electron orbit, then use a quantum assumption to get the radius, then solve for the energies.

Assumption 1: The atom is held together by the Coulomb Force

It is an experimental fact that the force F between two point charges q_1 and q_2 separated by a distance r is given by:

$$F = \frac{q_1 q_2}{4\pi \epsilon_0 r^2} \quad (\text{Eqn. 1.4a})$$

where $\frac{1}{4\pi \epsilon_0} = 8.99 \times 10^9 \left(\frac{\text{N m}^2}{\text{C}^2} \right)$ and q_1 and q_2 are in units of Coulombs. The distance, r , is in meters of course. Note that the charges may be positive or negative.

For a single electron atom, we take the charge of the nucleus, q_1 , to be $+Ze$, where Z is the atomic number of the atom (the number of protons in the nucleus). Z equals 1 for hydrogen. The charge of the electron, q_2 , is $-e$. Substituting the values into (1.4) above we obtain:

$$F = - \frac{Ze^2}{4\pi \epsilon_0 r^2} \quad (\text{Eqn. 1.4b})$$

The minus sign on the force term means the force is 'inward', or attractive.

Assumption 2: The electron moves in an elliptical orbit around the nucleus (as in planetary motion).

Let us assume that the electron moves in a circular orbit around the nucleus.

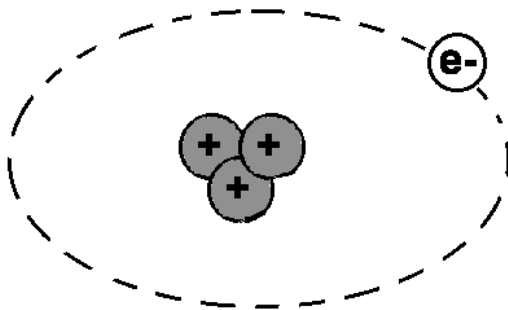


Figure 1.3 - Bohr atom model, $Z=3$

Then Newton's second Law ($F=ma$) (here: setting the Coulomb force equal to the centripetal force) can be written as:

$$\frac{-Ze^2}{4\pi \epsilon_0 r^2} = -mv^2 / r \quad (\text{Eqn. 1.5})$$

and we can now solve for the radius vs velocity.

Assumption 3: Quantized angular momentum

Bohr now introduced the first of his two new postulates, namely that the only allowed orbits were those for which the angular momentum, L , was given by:

$$L = mvr = n\hbar \quad (\text{Eqn. 1.6})$$

where:

m = electron mass; v = velocity; r = radius of the orbit; n = an integer (1,2,3,...); and

$$\hbar = \frac{h}{2\pi} = 1.054 \times 10^{-34} \text{ (Joule-seconds)} = 0.658 \times 10^{-15} \text{ (eV-seconds)}$$

where " \hbar " is simply Planck's constant, as before.

(One suggestion for a physical basis for this assumption is that if you view the electron as a wave, with wavelength $\lambda = h/p = h/mv$, then an integral number of wavelengths have to fit around the circumference defined by the orbit, or $n \cdot h/mv = 2\pi r$. Otherwise, the electron "interferes" with itself. This all follows as a corollary to the idea that an electromagnetic wave is a particle with energy $E = hf$ as above, and hence the momentum of a photon is $p = E/c = hf/c = h/\lambda$.)

This is sufficient to give us:

$$v_n = \frac{n\hbar}{mr_n} \quad (\text{Eqn. 1.7})$$

for the velocity of the electron in its orbit. Note that there is an index n , for the different allowed orbits. It follows that:

$$\frac{m v_n^2}{r_n} = \frac{Ze^2}{4\pi \epsilon_0 r_n^2} = \frac{mn^2 \hbar^2}{m^2 r_n^3} \quad (\text{Eqn. 1.8})$$

Upon solving for the radius of the orbit (r_n), we get:

$$r_n = \frac{n^2 \hbar^2}{m} \times \frac{4\pi \epsilon_0}{Ze^2} = n^2 \left(\frac{4\pi \epsilon_0 \hbar^2}{Zme^2} \right) \quad (\text{Eqn. 1.9})$$

$$r_n (\text{meters}) = n^2 \times 0.528 \times 10^{-10} / Z.$$

This only works for one electron atoms (H and He^+ as a practical matter), but within that restriction, it works fairly well. For hydrogen ($Z=1$) we get the Bohr radius, $r_1 = 0.528 \times 10^{-10}$ meters as the radius of the smallest orbit. The radius of the Bohr hydrogen atom is half an Angstrom. What is the radius of the orbit for the sole electron in He^+ (singly ionized helium, $Z=2$)?

Now we can solve for the energy levels.

The potential energy associated with the Coulomb force is:

$$U = \frac{q_1 q_2}{4\pi \epsilon_0 r} \quad (\text{Eqn. 1.10a})$$

taking $U(r = \infty) = 0$, and plugging in for the charges, we get

$$U = - \frac{Ze^2}{4\pi \epsilon_0 r} \quad (\text{Eqn. 1.10b})$$

A negative potential energy means the electron is in a potential 'well'. Given this expression for the potential energy, we need a similar expression for kinetic energy.

The kinetic energy, T , is easily obtained from equation 1.5:

$$T = \frac{1}{2} mv^2 = \frac{1}{2} \frac{Ze^2}{4\pi \epsilon_0 r} \quad (\text{Eqn. 1.11})$$

The total energy of the electron, E , therefore is obtained:

$$E = U + T = -\frac{Ze^2}{4\pi \epsilon_0 r} + \frac{1}{2} \frac{Ze^2}{(4\pi \epsilon_0) r} = -\frac{1}{2} \frac{Ze^2}{(4\pi \epsilon_0) r} \quad (\text{Eqn. 1.12})$$

The total energy is negative - a general characteristic of bound orbits. This equation also tells us that if we know the radius of the orbit (r) we can calculate the energy E of the electron.

Substituting the expression for r_n (Eqn. 1.9) into Eqn. 1.12, we obtain for the energy:

$$E = -\frac{1}{2} \frac{Ze^2}{4\pi \epsilon_0} \times \frac{1}{n^2} \frac{Zme^2}{4\pi \epsilon_0 \hbar^2}$$

or

$$E = -\frac{1}{2} \left(\frac{Ze^2}{4\pi \epsilon_0 \hbar} \right)^2 \frac{m}{n^2} = Z^2 \frac{E_1}{n^2} \quad (\text{Eqn. 1.13})$$

where

$$E_1 = - \frac{me^4}{32 \pi^2 \epsilon_0^2 \hbar^2} = - 13.58 \text{ eV}$$

is the energy of the electron in its lowest or "ground" state in the hydrogen atom.

Assumption 4: Radiation is emitted only from transitions between the discrete energy levels

The second Bohr postulate now defines the nature of the spectrum produced from these energy levels. This postulate declares that when an electron makes a transition from a higher to a lower energy level, a single photon will be emitted. This photon will have an energy equal to the difference in energy of the two levels. Similarly, a photon can only be absorbed if the energy of the photon corresponds to the difference in energy of the initial and final states. Schematically, we have:

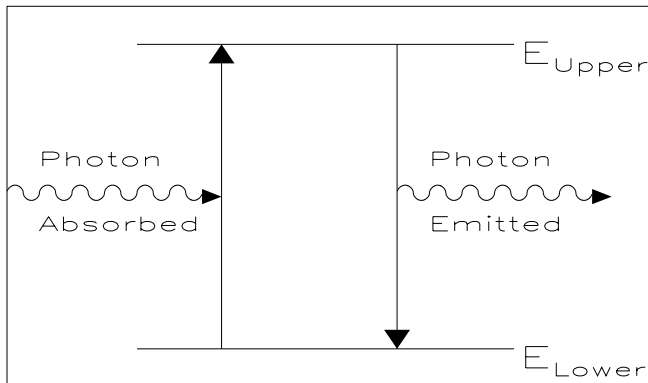


Figure 1.4 Second Bohr Postulate - photons produced/destroyed by discrete transition in energy.

Figure 1.5 illustrates the energy levels for the Bohr model of the hydrogen atom, and the associated "potential well" which is formed for the trapped orbits. We find that the ionization energy, the energy necessary to remove the electron from its "well", is 13.58 eV. If the electron gains somewhat less energy, it may move up to an "excited state", where $n > 1$.

For example, if it gains

$$13.58 - 3.39 = 10.19 \text{ eV}$$

it will move up to the $n = 2$ level.

Dropping down from $n = 2$ to $n = 1$, it will emit a photon of 10.19 eV energy, at a wavelength

$$\lambda = \frac{hc}{\Delta E} = 121.8 \text{ nm} = 1218 \text{ \AA}$$

If ΔE is expressed in electron-volts (eV), which it usually is, then the constant "hc" in the numerator can be written as:

$$hc = 4.14 \times 10^{-15} \cdot 3 \times 10^8 = 1.24 \times 10^{-6} \text{ (eV m)}$$

and thus the wavelength λ (in meters) is given by:

$$\lambda \text{ (m)} = \frac{1.24 \times 10^{-6}}{\Delta E \text{ (eV)}} \quad \text{or} \quad \lambda \text{ (nm)} = \frac{1240}{\Delta E \text{ (eV)}}$$

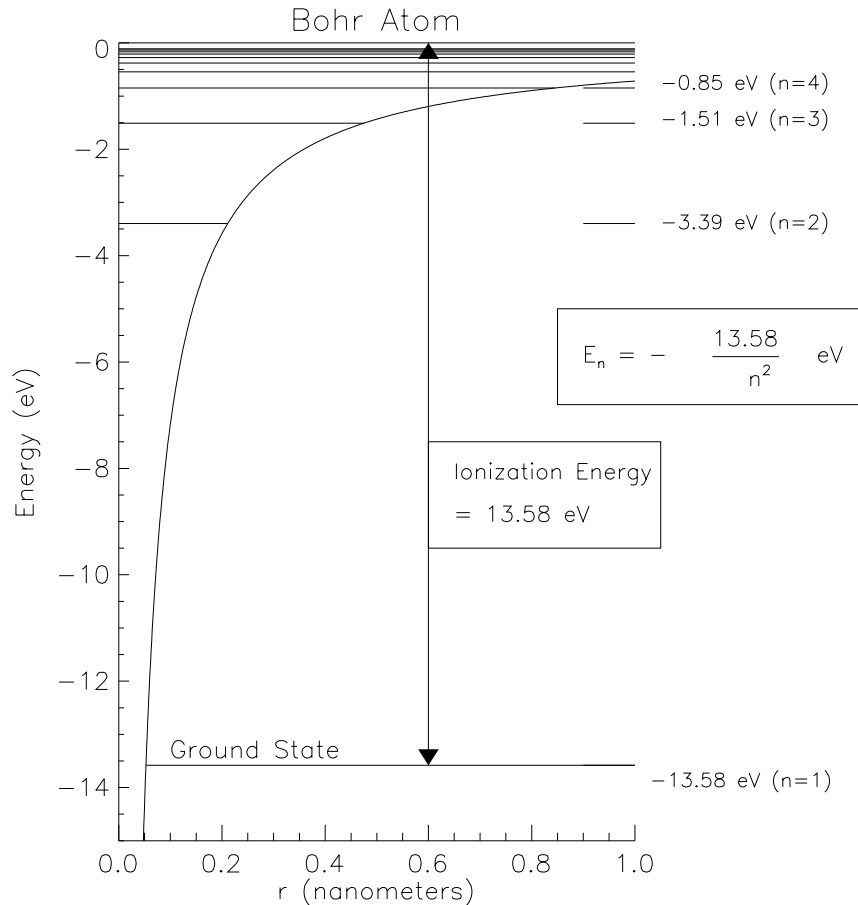


Figure 1.5 Energy levels for the hydrogen atom, according to the Bohr model.

In general, transitions will occur between different energy levels, resulting in a wide spectrum of discrete spectral lines. This is further illustrated in Figure 1.6. Transitions from (or to) the $n = 1$ energy level (the ground state) are called the Lyman series. The $n = 2$ to $n = 1$ transitions is the Lyman alpha (α) transition. This ultraviolet (UV) emission is one of the primary spectral (emission) lines of the sun's upper atmosphere. (See also, Figure 1.2) The emission (or absorption) lines in the visible portion of the sun's spectrum are the Balmer series, transitions from $n > 2$ to $n = 2$. Higher order series are of less importance for our purposes.

The Bohr model is successful in predicting the observed energy levels for one-electron atoms. It is useful for illustrating the quantum nature of the atom, and the associated energy levels. The model was ultimately replaced by the solution of the Schrodinger equation, and a more general form of quantum mechanics.

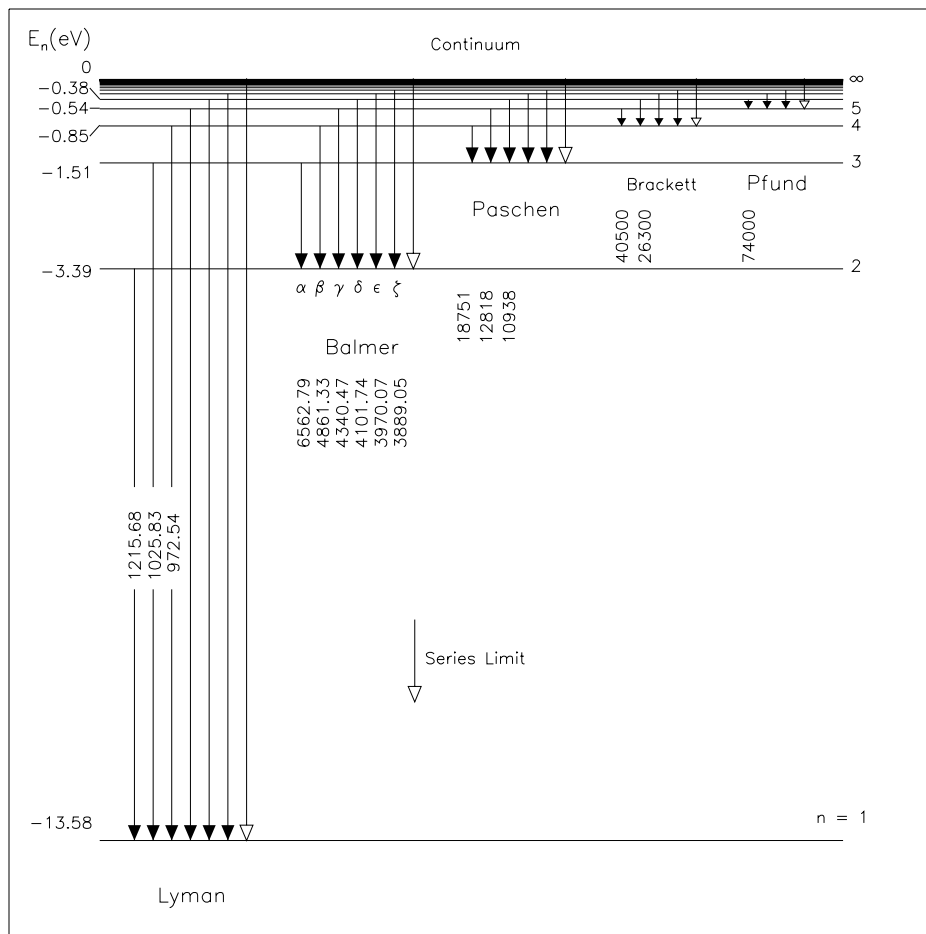


Figure 1.6. The energy level diagram of the hydrogen atom, showing the possible transitions corresponding to the different series. The numbers along the transitions are wavelengths. (Wavelengths are in units of Å, where 1 nm = 10 Å)

Adapted from : Fundamentals of Atomic Physics, Atam P. Arya, p264, 1971.

2 Black body radiation

Black Body Radiation is emitted by hot solids, liquids or dense gases and has a continuous distribution of radiated wavelength as shown in Figure 1.7. The shape and height of the curve in Figure 1.7 is determined by two important constants: the emissivity ϵ and the temperature T .

The emissivity of a surface is a measure of the efficiency with which the surface absorbs (or radiates) energy and lies between 0 (for a perfect reflector) and 1 (for a perfect absorber). A body which has $\epsilon = 1$ is called a black body and we shall assume that the radiators we deal with in space behave like black bodies.

Temperature is the most important parameter for black bodies and the amount of power radiated per unit surface area depends on temperature only. In Figure 1.7 we see a plot of radiated power/unit area/unit wavelength interval vs. wavelength known as Planck's Radiation Law. For our purposes we are particularly interested in two aspects: The total power radiated which is represented by the area under the curve and the wavelength at which the curve peaks, λ_{\max} .

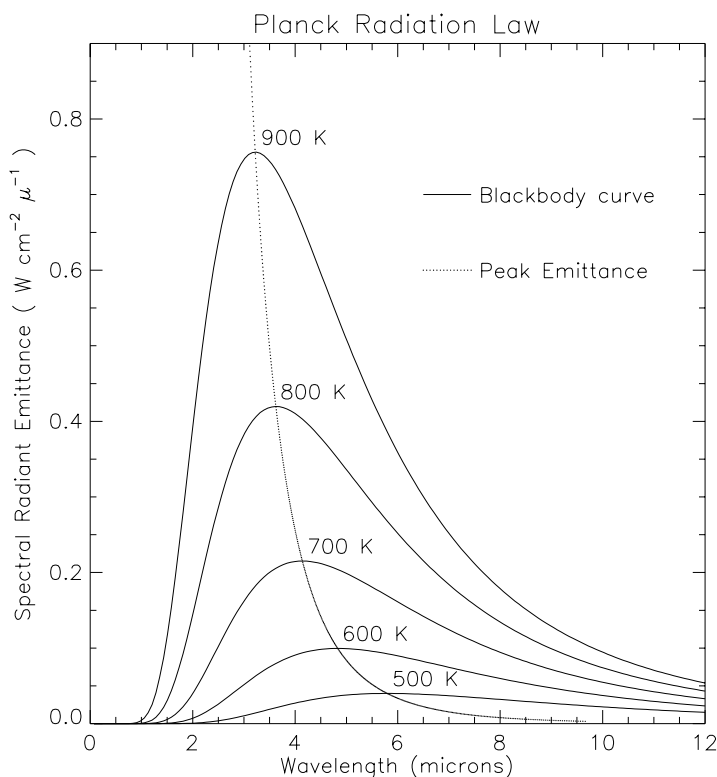


Figure 1.7 Black body radiation spectrum

$$U(\lambda) = \frac{8\pi hc}{\lambda^5} \frac{1}{e^{\left(\frac{hc}{\lambda kT}\right)} - 1}$$

$$= \frac{8\pi (kT)^5}{(hc)^4} \frac{\left(\frac{hc}{\lambda kT}\right)^5}{e^{\left(\frac{hc}{\lambda kT}\right)} - 1}$$

Student exercise: check long wavelength behavior - use the fact that for small x :

$$e^x - 1 \approx x$$

to get rid of the exponential term in the denominator.

Also, which term dominates U for small wavelength ?

The power radiated (integrated over all wavelengths) is given by

$$R = \sigma \varepsilon T^4 \left(\frac{\text{Watts}}{\text{m}^2} \right) \quad (\text{Stefan Boltzmann Law}) \quad (\text{Eqn. 1.14})$$

where R = Power radiated / m^2

ε = Emissivity (taken as unity for black body)

$$\sigma = 5.67 \times 10^{-8} \left(\frac{\text{W}}{\text{m}^2 \text{ K}^4} \right) \quad (\text{Stefan's Constant})$$

T = Temperature of the radiator (in K)

The wavelength at which the peak in radiation occurs is given by Wien's Displacement Law:

$$\lambda_{\max} = \frac{a}{T} \quad (\text{Eqn. 1.15})$$

for a given temperature T . The constant "a" has the value

$$a = 2.898 \times 10^{-3} \text{ (m K)}$$

which gives λ_{\max} in meters if T is K.

Example:

Assume that the sun radiates like a blackbody, which is not a bad assumption, though we must choose two slightly different temperatures to match the observed quantities.

(a) Find the wavelength at which this radiation peaks, λ_{\max} . The solar spectral shape in the visible is best matched by a temperature of ~ 6000 K.

(b) Find the total power radiated by the sun. The Stefan-Boltzmann law is best served by an "effective temperature" of ~ 5800 K.

Solution:

$$\lambda_{\max} = \frac{a}{T} = \frac{2.898 \times 10^{-3} \text{ (m / K)}}{6000 \text{ K}} = 4.83 \times 10^{-7} \text{ m}$$

The spectrum peaks at ~ 500 nm, as illustrated below in Figure 1.8.

Next we can calculate R , the power emitted per-square-meter of surface. We use: $R = \sigma \varepsilon T^4$ and we assume that $\varepsilon = 1$ (Black Body). Evaluating, we get:

$$R = 5.67 \times 10^{-8} \bullet 1 \bullet 5800^4 = 6.42 \times 10^7 \frac{\text{Watts}}{\text{meter}^2}$$

To find the total solar power output we must multiply by the solar surface area, $S_{\odot} = 4 \pi R_{\odot}^2$,

where $R_{\odot} = 6.96 \times 10^8$ m is the mean radius of the sun. Hence the total solar power output is:

$$P_{\odot} = R (4 \pi R_{\odot}^2) = 4 \pi (6.96 \times 10^8)^2 \times (6.42 \times 10^7)$$

$$P_{\odot} = 3.91 \times 10^{26} \text{ W}$$

See Kenneth Phillips, *Guide to the Sun*, Cambridge Press, 1992., pages 83-84) The sun's spectrum is shown in Figure 1.8, with the spectrum of a 5,800 K black body superimposed.

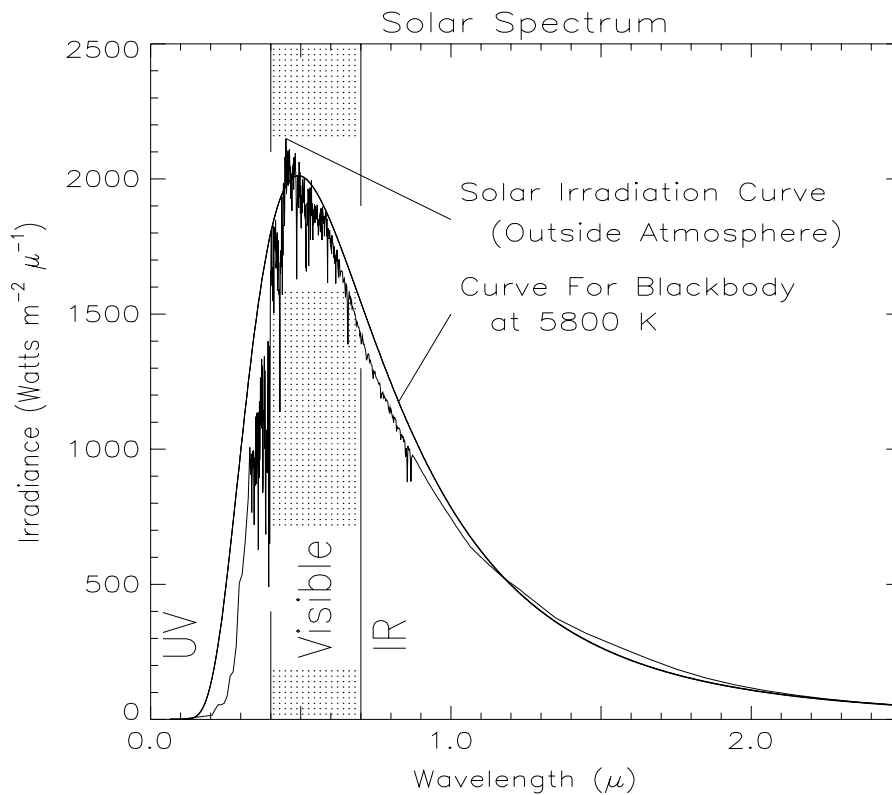


Figure 1.8 - The solar spectrum, based on the spectrum of Neckel and Labs "The solar radiation between 3300 and 12500 Angstrom, *Solar Physics*, 90, 205-258, 1984. Data file courtesy of Bo-Cai Gao, NRL. The peak occurs at about 460 nm (blue).

3 X-rays - Brehmstrahlung

An important source of x-rays in space comes from the sudden deceleration of high speed electrons. This is also a common process by which x-rays are generated in the laboratory (or the dentist's office). It is generally true that when an electron is slowed down by collisions or deflected by an electric or magnetic field it will radiate electromagnetic energy (photons). The production mechanism most commonly found in space involves a (mono-)energetic beam of electrons which are decelerated by impact with a material surface or a dense gas (or plasma). The electron beam will produce radiation with a continuous spectrum of frequencies up to a maximum frequency, f_{\max} , given by: $E = hf_{\max}$. Here, E is the original electron energy, or the energy of the electron beam in the case being presented here. Figure 1.9 shows the type of spectrum produced in the laboratory, in this case exposing a tungsten target to electron beams at various energies. Note that in this figure, the spectra are plotted vs. wavelength, not frequency. Hence, the cutoff at f_{\max} corresponds to a spectrum which terminates at a minimum wavelength, λ_{\min} , below which no x-rays are generated.

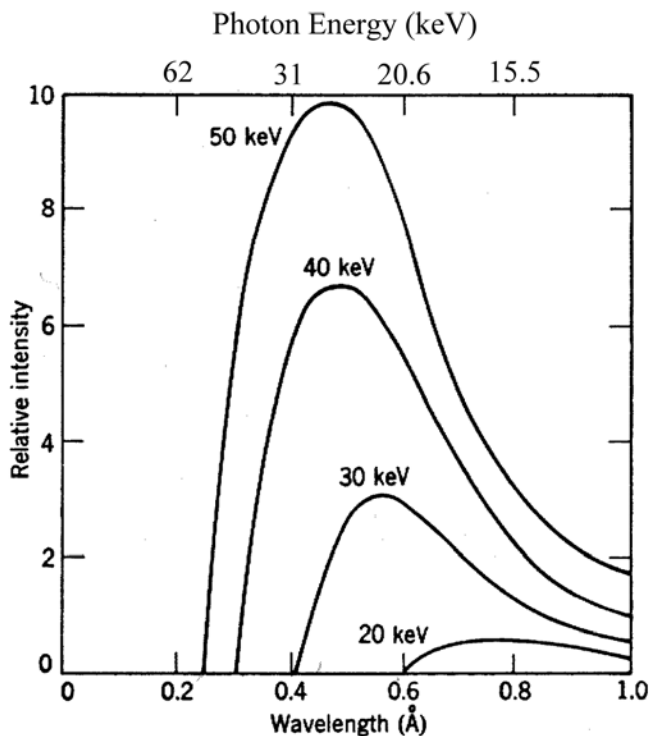


Figure 1.9 Spectrum for x-rays generated by an electron beam (Brehmstrahlung). From: Eisberg and Resnick, Quantum Physics, (fig 2-10), 1985.

C Kinetic Theory

Before proceeding into the concepts of plasma physics, a few elements of the classical kinetic theory of gases need to be established - primarily the relationship between the temperature of an ensemble (collection) of atoms, and the kinetic energy of individual atoms.

1 Temperature and Kinetic Energy

Kinetic theory says that the mean kinetic energy of a molecule in a gas is related to the temperature of the gas by the relation:

$$\frac{1}{2} m v_{\text{rms}}^2 = \frac{3}{2} kT \quad (\text{Eqn. 1.16})$$

where $v_{\text{rms}} = \sqrt{\langle v^2 \rangle}$.

Here, k is the Boltzmann constant, $k = 1.38 \times 10^{-23} \text{ J/K}$, and the brackets $\langle \rangle$ indicate an average over the volume of particles at hand. The velocity (v_{rms}) is the root-mean-square velocity. The energy described by this relation is again the average kinetic energy of an atom or molecule. These particles are colliding with each other continually, and these collisions can excite electronic transitions in the atoms or molecules. If the collisions are sufficiently energetic, an electron can be knocked free, ionizing the particle. For a hydrogen atom, the energy required is $E \approx 13.6 \text{ eV}$. The temperature associated with this (average kinetic) energy is therefore:

$$kT \approx 13.6 \text{ eV} \cdot 1.6 \times 10^{-19} \left(\frac{\text{J}}{\text{eV}} \right) \quad \text{or} \quad T \approx \frac{13.6 \cdot 1.6 \times 10^{-19}}{1.38 \times 10^{-23}} = 1.6 \times 10^5 \text{ K}$$

We can see that the temperature required for "thermal" ionization of hydrogen atoms is of such a magnitude as is normally only found in regions of very high temperatures such as the solar corona. (Note that the factor of $\frac{3}{2}$ has been cheerfully ignored - partly to avoid a false implication of accuracy in the above estimate).

2 Maxwellian Distributions

In equilibrium distributions of gas molecules, where concepts such as temperature make sense, it is typically found (and often assumed) that the distribution of atomic or molecular velocities follow the well-known Maxwellian distribution. This will occasionally be referred to in the following material, and it is helpful to illustrate the distribution here.

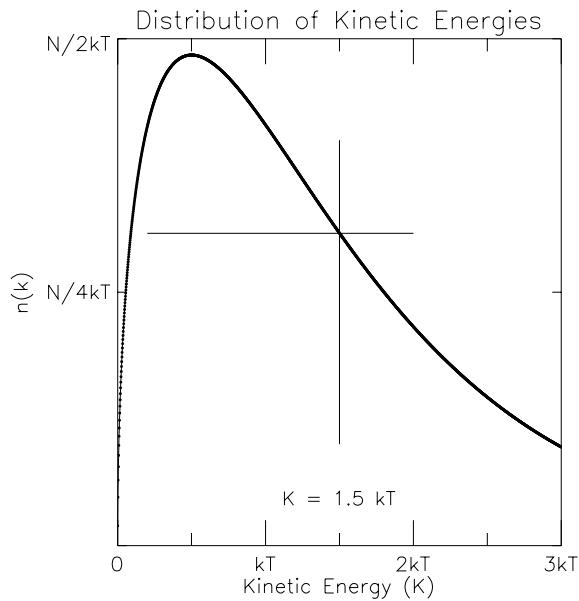


Figure 1.10 Maxwellian Distribution

$$n(K) = \frac{2N \sqrt{K} e^{-K/kT}}{\sqrt{\pi}(kT)^{3/2}} \quad (\text{Eqn. 1.17})$$

gives the number of particles, n , which have a kinetic energy, K , for a gas volume with a total number of particles (e.g. molecules), N , at a temperature, T . Here, this equation is used in the well defined, and classic, integral:

$$N = \int_{K=0}^{\infty} \frac{2N \sqrt{K} e^{-K/kT}}{\sqrt{\pi}(kT)^{3/2}} dK$$

in order to determine the total number of particles, N , in a given volume. Useful questions: Where does the distribution peak? What is the value of $n(K)$ at the maximum? What is the value at $K = 1.5 kT$?

D Plasmas

1 Phases of matter

It is estimated that most of the matter in the universe exists in the plasma state and that solid objects such as the planets are only minor specks of dust floating in this plasma space. Plasmas are distinguished from the traditional three states of matter primarily by the temperature range at which most plasmas are found. Figure 1.11 illustrates the range of temperatures for which matter generally exists in a plasma state.

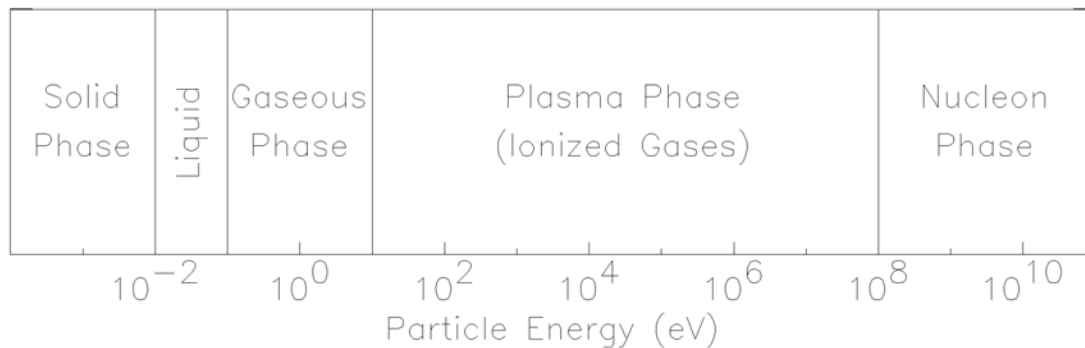


Figure 1.11 Phases of matter as a function of average particle energy (temperature)

A plasma is a quasi-neutral gas of charged and neutral particles which exhibits collective behavior. The charged particles are typically (positively charged) ions and electrons, and the net charge in a region is typically very small. Plasmas typically consist of only partially ionized gas. For example, the gas found at 1-5 earth radii (altitude) is ~50% ionized. The definition of a plasma can be quantified, to an extent, by the usage of two quantities defined here, the Debye Length (λ_D) and the plasma frequency (ω_p). We begin by developing the Debye Length.

2 Plasma Parameters

Debye Shielding length (λ_D)

One important characteristic of a plasma is the shielding of electric fields produced by the ensemble of particles. The concept of the Debye length (λ_D) also allows us to help define when we have a plasma.

Conceptually, we begin with an object in a plasma (e.g. a probe, or satellite) with a charge on it. In a vacuum we would have the Coulomb potential

$$\phi(r) = \frac{Q}{4\pi\epsilon_0} \frac{1}{r} \quad (\text{Eqn. 1.18})$$

If we immerse this same point charge +Q in a plasma it will attract some of the negative charges and repel the positive charges leading to a partial shielding of the charge at distant points. The potential at a point distant r from our charge +Q now becomes:

$$\phi(r) = \frac{Q}{4\pi\epsilon_0} \frac{1}{r} e^{-r/\lambda_D} \quad (\text{Eqn. 1.19})$$

where λ_D (meters) = $\left(\frac{\epsilon_0 kT}{e^2 n_e} \right)^{1/2}$ (Eqn. 1.20)

is the Debye length, and

n_e = electron density (m^{-3})

k = Boltzmann constant = 1.381×10^{-23} (J/ K)

T = Temperature (K)

e = electronic charge = 1.602×10^{-19} (Coulombs)

ϵ_0 = 8.854×10^{-12} (Farads/m)

The two important parameters are the temperature of the electrons T (or equivalently their average kinetic energy kT) and the electron density n_e . For quick calculations we may write the Debye length (λ_D) as:

$$\lambda_D (m) = 69 \left(\frac{T}{n_e} \right)^{1/2} \text{ where } T \text{ is in K, } n_e \text{ is in electrons/m}^3 \quad (\text{Eqn. 1.21})$$

$$\lambda_D (m) = 7430 \left(\frac{kT}{n_e} \right)^{1/2} \text{ where } kT \text{ is in eV, } n_e \text{ is in electrons/m}^3 \quad (\text{Eqn. 1.22})$$

Note the inverse dependence on density - higher density plasmas provide more shielding. (Also note, it is common in space physics to use electron Volts (eV) as a unit of temperature, as well as a non-mks unit of energy.)

Example:

$n = 5 \times 10^{10}$ electrons/ m^3 ; $T = 22,000$ K	$n = 5 \times 10^4$ electrons/ cm^3 , (Note units - must convert to electrons/ m^3); $T = 2$ eV
$\lambda_D (m) = \left(\frac{\epsilon_0 kT}{e^2 n_e} \right)^{1/2}$ $= \left(\frac{8.85 \times 10^{-12} \cdot 1.38 \times 10^{-23} \cdot 22 \times 10^3}{(1.6 \times 10^{-19})^2 \cdot 5 \times 10^{10}} \right)^{1/2}$ $= 0.0458 \text{ meters} = 4.6 \text{ cm}$	$\lambda_D (m) = \left(\frac{\epsilon_0 kT}{e^2 n_e} \right)^{1/2}$ $= \left(\frac{8.85 \times 10^{-12} \cdot 1.6 \times 10^{-19} \cdot 2}{(1.6 \times 10^{-19})^2 \cdot (5 \times 10^4 \cdot 10^6)} \right)^{1/2}$ $= 0.047 \text{ meters} = 4.7 \text{ cm}$
$\lambda_D (m) = 69 \left(\frac{T}{n_e} \right)^{1/2} = 69 \left(\frac{22 \times 10^3}{5 \times 10^{10}} \right)^{1/2}$ $= 0.0458 \text{ meter} = 4.6 \text{ cm}$	$\lambda_D (m) = 7430 \left(\frac{kT}{n_e} \right)^{1/2} = 7430 \left(\frac{2}{5 \times 10^4 \cdot 10^6} \right)^{1/2}$ $= 0.047 \text{ meters} = 4.7 \text{ cm}$

There is a reason why the answers in the two columns are (almost) the same - the 22,000 K temperature converts to 2 eV.

Plasma Frequency (ω_p)

The collective behavior of plasmas is, in many ways, exhibited most strongly in the enormous variety of electromagnetic waves which occur in ionized media. The most prominent of these wave modes is the plasma oscillation, which is a result of the oscillation of the electrons about the (relatively massive and immobile) ions. Much as in the case of a mass on a spring, the mass of the electrons determines the characteristic frequency. The total mass (per unit volume) is of course proportional to the electron density. The plasma frequency is defined as:

$$\omega_p = \left(\frac{n_e e^2}{\epsilon_0 m_e} \right)^{\frac{1}{2}} \quad (\text{Eqn. 1.23})$$

If the various constants are plugged in, we obtain:

$$\omega_p \text{ (radians/second)} = 56 \sqrt{n_e \text{ (electrons/meter}^3\text{)}}.$$

In Figure 1.12 the ranges of several plasmas are shown as functions of the value of the Debye shielding length. Recall that the conversion from frequency in radians/second to Hertz is given by:

$$f_p \text{ (Hz)} = \frac{\omega_p \text{ (radians/second)}}{2\pi}$$

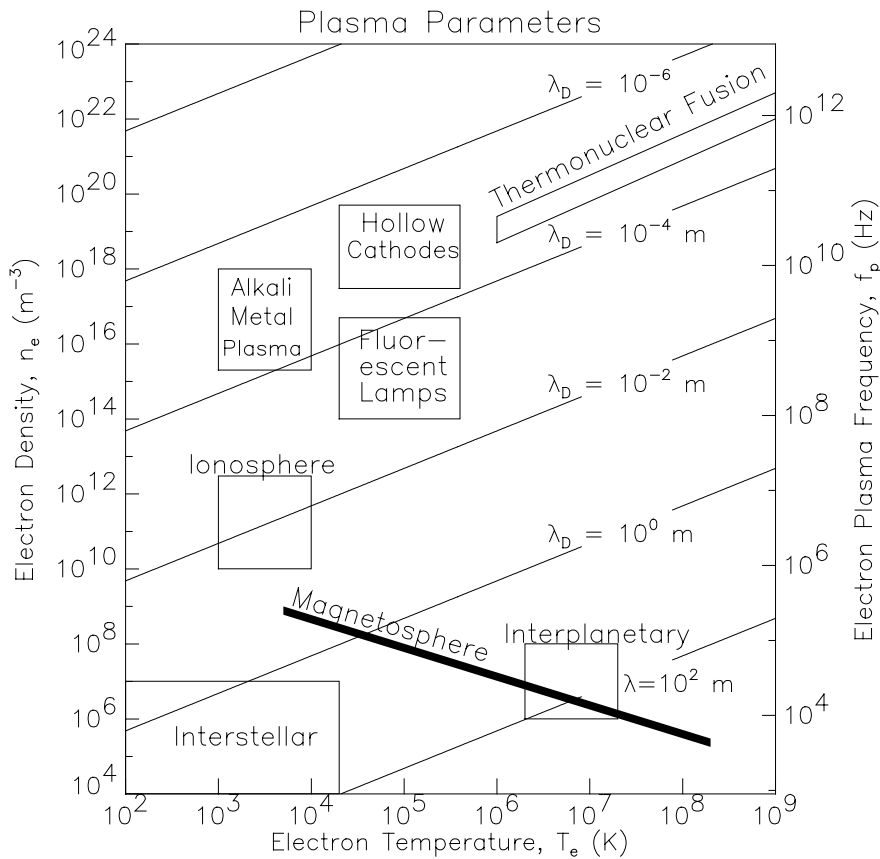


Figure 1.12 - Plasma parameters and Debye shielding length

3 Definition of Plasma

With the above two quantities defined, we are now able to consider the criteria which define the existence of a plasma.

1. The typical linear dimensions of the plasma (e.g. L) must be much larger than the characteristic plasma dimension, λ_D , the distance scale over which electric fields penetrate into a plasma, e.g.:

$$\lambda_D \ll L \quad (\text{Eqn. 1.24})$$

The characteristic scale size for variations in the ionosphere or higher altitude space regimes is many kilometers, and we see that the first criterion is easily met for the ionized particle distributions found in space.

2. The number of charges (N_D) in a sphere of radius λ_D must be very much larger than 1,

$$N_D = \left(\frac{4}{3} \pi \lambda_D^3 \right) (n_e) = 1.38 \times 10^6 \frac{T^{3/2}}{n_e^{1/2}} \gg 1 \quad (\text{Eqn. 1.25}).$$

3. The plasma frequency must be larger than the frequency of collisions. This can be written as $\omega_p \gg \nu_c$, where ν_c is the collision frequency. If this criterion is not met, there may be a high degree of ionization, but collective (wave) phenomena cannot occur.

Example:

Consider whether the terrestrial ionosphere can be considered to be a plasma by our criteria.

First condition: Typical experimental data for the ionosphere at an altitude of 90 km, mid-latitude, daylight are: $T \sim 800^\circ \text{ K}$ and $n_e = 10^{11} \text{ electrons / m}^3$

$$\lambda_D = 69 \left(\frac{T}{n} \right)^{1/2} = 6 \times 10^{-3} \text{ m} = 6 \text{ mm}$$

Since the typical dimensions of the ionosphere are on the order of 100 km or more, the condition that the system be much larger than the Debye length is certainly satisfied

Second condition: $N_D \gg 1$.

$$N_D = 1.38 \times 10^6 \frac{T^{3/2}}{n_e^{1/2}} = 1.38 \times 10^6 \frac{800^{3/2}}{(10^{11})^{1/2}}$$

$$N_D = 9.87 \times 10^4 \approx 10^5$$

so the second condition is also satisfied.

Third condition: This criterion requires the additional piece of information, that of the collision frequency (ν_c) at 90 km, which has a nominal value of 1.3×10^5 Hz. Comparing to the plasma frequency, we find $\omega_p = 56\sqrt{8.2 \times 10^9} = 5.1 \times 10^6$ (radians/second). The third criterion is satisfied reasonably well, but in fact collisional processes need to be considered carefully when studying ionospheric processes. Collisional frequencies drop rapidly with increasing altitude.

4 Plasma Theories and Approximations

The complete description of a plasma must include the motion of charged particles in time dependent electric and magnetic fields both applied and self generated. Kinetic Theory and the Boltzmann Transport Equation can be used to formulate this problem, but in practice the solutions are so difficult that little use is made of these equations for the actual solutions of plasma problems.

There exists however two approximations which apply in limiting situations and which are more manageable:

(a) The magneto-hydrodynamic approximation: When the particle density is high enough and collisions between the particles are frequent enough that we may define a thermodynamic temperature and pressure then we may treat the plasma as a conducting fluid. A combination of Maxwell's and fluid dynamic equations can be used to calculate such effects as plasma oscillations, hydromagnetic waves, etc. These equations are still quite complex but with certain approximations they can be solved. The concept of Debye shielding, described briefly above, arises from the hydrodynamic approximation.

(b) Single Particle Limit: When collisions between particles become very infrequent we can use the other limiting approximation, namely orbit theory. Here we assume that individual charges move in orbits determined by the applied electric and magnetic fields and that the fields generated by the moving charges themselves are small and can be neglected.

Both of these approaches are used in the analysis of plasma phenomena in space. We shall start with a brief review of electric and magnetic fields, and the forces they exert on electric charges. The single particle limit (orbit theory) is based directly on these concepts.

5 Orbit theory - Motion of Charged Particles

Lorentz Force

The electric field \vec{E} is defined as the force exerted on a small positive test charge q_0 placed at the point in question. This is usually written as

$$\vec{E} = \lim_{q_0 \rightarrow 0} \frac{\vec{F}_E}{q_0} \quad (\text{Eqn. 1.26})$$

(q_0 is positive and small enough so its effect on the other charges is negligible)

The units of electric field are: $\frac{\text{Newtons}}{\text{Coulomb}} = \frac{\text{Volts}}{\text{meter}}$

The magnetic field of induction \vec{B} is defined by magnetic force \vec{F}_B which a particle carrying a charge q will experience when it moves with velocity \vec{v} at the point in question. Thus the defining equation for \vec{B} may be written as

$$\vec{F}_B = q (\vec{v} \times \vec{B}) \quad (\text{Eqn. 1.27})$$

The unit of \vec{B} is the Tesla, obtained directly from this equation.

Thus the force experienced by a charged particle moving in a combined electric (\vec{E}) and magnetic (\vec{B}) field is the so called Lorentz force given by

$$\vec{F} = q (\vec{E} + \vec{v} \times \vec{B}) \quad (\text{Eqn. 1.28})$$

Motion of Charged Particles in Simple \vec{E} and \vec{B} Fields

Let us consider a couple of simple examples of a particle of constant mass m and charge q (can be + or -) moving under the influence of electric or magnetic forces. The actual motions can then (often) be understood as a combination of the motions defined below.

Example 1: $B=0$, Uniform $E \neq 0$

Assume $\vec{B} = 0$ and \vec{E} is uniform in space and constant in time.

The force on our particle is:

$$\vec{F} = q\vec{E}$$

and from Newton's Second Law

$$\vec{F} = m\vec{a} = q\vec{E} \Rightarrow \vec{a} = \frac{q}{m}\vec{E} = \text{a constant}$$

We thus have motion exactly analogous to motion under constant gravitational acceleration \vec{g} but with one major difference:

These accelerations due to the \vec{E} field can be along \vec{E} (if $q > 0$), or along $-\vec{E}$ (if $q < 0$). Thus electrons and negative ions will accelerate in the $-\vec{E}$ direction, while positive ions will accelerate in the direction of \vec{E} .

Example 2: $E=0$, Uniform $B \neq 0$

Assume $\vec{E} = 0$ and \vec{B} is uniform in space and constant in time. The motion will now depend on the direction of the initial velocity \vec{v}_0 .

Special case 1: ($\vec{v}_0 \parallel \vec{B}$)

The initial velocity vector is parallel to the magnetic field. ($\vec{v}_0 \parallel \vec{B}$)

$$\vec{F}_B = q(\vec{v} \times \vec{B}) = 0$$

The particle will continue with its initial velocity \vec{v}_0 .

Special case 2: ($\vec{v}_0 \perp \vec{B}$)

The initial velocity vector is perpendicular to the magnetic field. ($\vec{v}_0 \perp \vec{B}$). The magnitude of the deflecting force is now a constant $F_B = qv_0 B$ and always perpendicular to the instantaneous velocity vector, whose magnitude is also constant. This force will result in circular motion with radius r_c given by

$$r_c = \frac{mv_0}{qB} \quad (\text{Eqn. 1.29})$$

This quantity is known as the cyclotron or gyration radius and plays an important role in much of the material to follow. Equally important are the period T of the motion and its inverse, the gyrofrequency.

$$T = \frac{2\pi r_c}{v_0} = \frac{2\pi mv_0}{qB v_0} = \frac{2\pi m}{qB} \quad (\text{seconds})$$

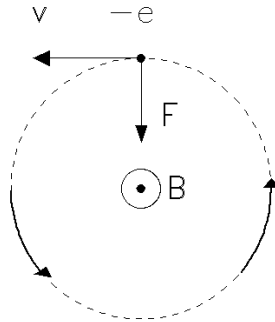
or

$$f_c = \frac{1}{T} = \frac{qB}{2\pi m} \quad (\text{revolutions/second}) \Rightarrow \quad (\text{Eqn. 1.30})$$

$$\omega_c = 2\pi f_c = \frac{qB}{m} \quad (\text{radians/second})$$

The gyrofrequency is independent of the speed and hence the energy of the gyrating particle. In a given field it depends only on the charge to mass ratio of the particle. Consider a uniform field B pointing out of the paper. Particles of opposite charge will gyrate with opposite direction as shown. Note that the particle motion is diamagnetic - the magnetic field created by the orbiting charged particle opposes the applied field.

Negative Charge



Positive Charge

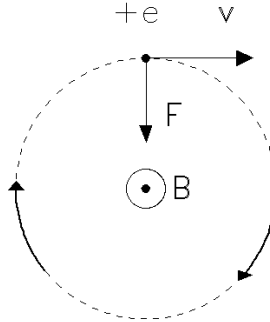


Figure 1.13 Charged particle motion in a uniform magnetic field. The magnetic field is 'up', out of the page.

General Case: Arbitrary direction of \vec{v}_0

If the initial velocity vector is neither parallel nor perpendicular to the \vec{B} field the trajectory will be a helix of constant pitch and radius which results from the superposition of circular motion in the perpendicular plane and uniform translational motion along the field direction. Hence the axis of the helix will be parallel to \vec{B} .

Particle Drifts: $E \neq 0$, $B \neq 0$

In the first two examples above, one of the two terms (E , B) is zero. If both are non-zero, a peculiar result is obtained. It is found that for any force \vec{F} applied to a particle, in combination with the magnetic force, the equation of motion produces a motion for the center of the orbit termed a "drift velocity" given by:

$$\vec{v}_D = \frac{1}{q} \frac{\vec{F}_\perp \times \vec{B}}{B^2}. \quad (\text{Eqn. 1.31})$$

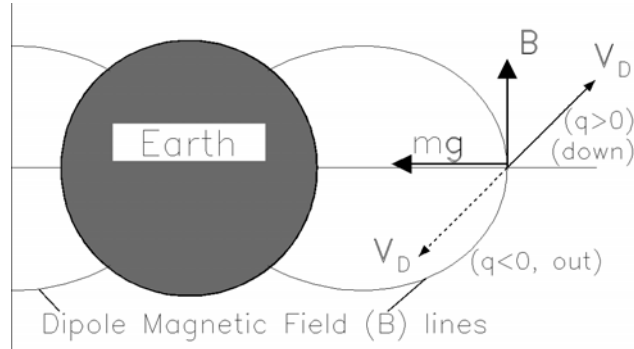
\vec{F}_\perp is the component of the force perpendicular to the direction of \vec{B} . (Realistically, this is almost always the electrical force, $F = qE$) We shall now consider two examples of Eqn. 1.31:

Example 1:

Suppose the additional force \vec{F} is the gravitational force $m\vec{g}$ and let us assume that it is perpendicular to \vec{B} (This would be the case at the geomagnetic equator). Then:

$$\vec{V}_D = \frac{1}{q} \frac{m\vec{g} \times \vec{B}}{B^2}$$

Figure 1.14 Drift directions due to gravity. Positively charged particles drift "down" or into the page, negatively charged particles drift "up" or out of the page.



It can be seen that positive and negative charges will drift in opposite directions. Thus protons will precess in an easterly directions and electrons in a westerly direction forming ring currents around the earth. The magnitude of the drift velocities for this case are rather small due to the small mass of both the proton and electron.

At an altitude of one earth radius (6400 km); the geomagnetic field has a magnitude of 3.75×10^{-6} T (3.75 μ T). The charge of both proton and electron has a magnitude of 1.6×10^{-19} Coulombs. The proton mass is 1.67×10^{-27} kg and the electron mass is 9.1×10^{-31} kg.

Putting all these numbers into the formula we obtain

Proton drift velocity $\approx 3 \times 10^{-2}$ m/sec

Electron drift velocity $\approx 1.5 \times 10^{-5}$ m/sec

Note however that positive charges moving eastward and negative charges moving westward both contribute to a net eastward current, so that a net current is generated.

Example 2:

Suppose that we now place our moving charge in a crossed electric and magnetic

field so that our force \vec{F} in equation 1.31 is given by: $\vec{F} = q\vec{E}$

and we again assume that \vec{E} (and hence \vec{F}) is perpendicular to \vec{B} as shown below

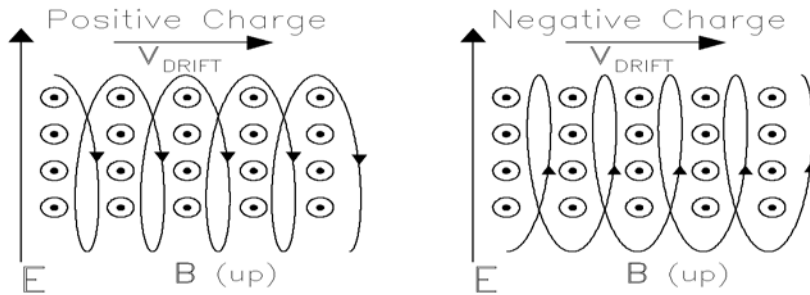


Figure 1.15 Drift directions due to an electric field

$$\vec{v}_D = \frac{1}{q} \frac{q \vec{E} \times \vec{B}}{B^2} = \frac{\vec{E} \times \vec{B}}{B^2} \quad (\text{Eqn. 1.32})$$

Which says that both positive and negative charges drift in the same direction (to the right in our case). This happens because the force \vec{F} is charge dependent and q cancels in the equation for \vec{v}_D . Hence there is no net current in this case.

We can see why the drift is to the right in both cases. For positive charges (clockwise rotation) the energy (speed) is lower at the bottom and hence the radius of curvature is smaller. Near the top the particle has picked up energy from the electric field and hence its radius of curvature is now larger. The net effect is a displacement to the right.

For the negative charge the sense of rotation is reversed and the electric force acting on it is now downward. Combining these two effects again produces a drift toward the right as shown in the sketch. Note that the electric field induced drift does not result in a net current.

In our study of particle motion in the geomagnetic field we shall consider two more cases of particle drifts caused by gradients and curvature in the field itself.

A brief derivation of the drift equation

The drift equation can be derived in a fairly simple form for the Lorentz force. We make a few simplifying choices about directions, and then proceed. The magnetic field is taken to be in the z direction, the electric field in the y direction.

To begin:

$$\vec{F} = q(E_y \hat{y} + \vec{v} \times B_z \hat{z}) \quad (\text{Eqn. 1.28})$$

which has two (interesting) components, F_x and F_y .

$$\begin{aligned} F_x &= m\ddot{x} = 0 + q v_y B_z \\ F_y &= m\ddot{y} = qE_y - q v_x B_z \end{aligned} \quad (\text{Eqn. 1.33})$$

Note that the acceleration is just the derivative of the velocity:

$$\begin{aligned} m\dot{v}_x &= q v_y B_z \\ m\dot{v}_y &= qE_y - q v_x B_z \end{aligned} \quad (\text{Eqn. 1.34})$$

Now we know that in the absence of an electric field, the solutions to these two equations is just circular motion, the primary motion of a charged particle in a magnetic field.

$$E = 0 \Rightarrow v_x = v_o \cos \omega t; v_y = \mp v_o \sin \omega t \quad (\text{Eqn. 1.35})$$

where the top sign is for positively charged particles (protons), the lower for negatively charged particles (electrons). In fact, by being just a bit clever, we can make use of these solutions to solve the current problem. The solution comes by forming a new variable, v'_x which is comprised of the circular motion defined above, and a constant offset. Note that this constant offset has a derivative with respect to time that is zero.

$$v'_x = v_x - v_o; \dot{v}'_x = \dot{v}_x \quad (\text{Eqn. 1.36})$$

now we use this new variable in the above differential equation:

$$\begin{aligned} m\dot{v}'_x &= q v_y B_z \\ m\dot{v}_y &= qE_y - q(v'_x + v_o) B_z \end{aligned} \quad (\text{Eqn. 1.37})$$

if $E_y = v_o B_z$ or $v_o = \frac{E_y}{B_z}$ the new equations now reduce to:

$$\begin{aligned} m\dot{v}'_x &= q v_y B_z \\ m\dot{v}_y &= -q v'_x B_z \end{aligned} \quad (\text{Eqn. 1.38})$$

which are already known to have a solution which is simply uniform circular motion (Exercise for the student: show that the solutions given in Eqn. 1.35 solve Eqn. 1.38). Hence, v_x is simply

$$v_x = \mp v_o \sin \omega t + \frac{E_y}{B_z} \quad (\text{Eqn. 1.39})$$

This can be compared to Eqn. 1.32; the constant term is the drift velocity, $\vec{v}_D = \frac{\vec{E} \times \vec{B}}{B^2}$

6 A few results from the magneto-hydrodynamic approximation

Magnetic Pressure

We shall now use the fluid approximation in which we assume that a plasma can be considered to be an electrically conducting fluid which moves under the influence of electrical, magnetic, gravitational and pressure gradient forces. Consider a streaming plasma which has a magnetic field B embedded within it. If the plasma is not accelerating or being constricted in its flow we can show that

$$P_K + \frac{B^2}{2\mu_0} = \text{constant} \quad (\text{Eqn. 1.40})$$

where P_K = pressure due to particle motion (N/m^2)
(usually equal to the plasma pressure, nkT)

$$\begin{aligned} \frac{B^2}{2\mu_0} &= \text{Magnetic energy density} \left(\frac{\text{J}}{\text{m}^3} \right) \\ &= \text{Magnetic pressure} = P_M \end{aligned}$$

and $\mu_0 = 4\pi \times 10^{-7} \text{ (H/m)}$, the permeability constant.

This equation says that in a region in which there is both plasma and magnetic field present we can consider the magnetic energy density to play the role of a magnetic pressure P_M . (Dimensionally energy density and pressure are equivalent). The fact that the sum of kinetic pressure P_K and the magnetic pressure P_M is constant turns out to be of particular interest at the interface between plasma regions and field regions.

In regions where both stationary plasmas and magnetic fields are present it is customary to define a parameter β given by

$$\beta = \frac{\text{Kinetic Energy Density}}{\text{Magnetic Energy Density}} = \frac{\sum_i n_i kT_i}{B^2 / 2\mu_0} \quad (\text{Eqn. 1.41})$$

where n_i = particle density i_{th} species (m^{-3})
 k = Boltzmann's Constant
 T_i = Temperature of i_{th} species

Thus the numerator represents the kinetic energy density of the plasma and the denominator represents the magnetic energy density. The numerator is typically equal to $n(kT_{\text{electron}} + kT_{\text{ion}})$, with the ions and electrons having a common density (quasi-neutrality), but not necessarily a common temperature.

Thus a high β plasma has a large kinetic energy density and the behavior of the system will be dominated by the particles. If on the other hand β is very small the magnetic field dominates and determines the behavior of the system.

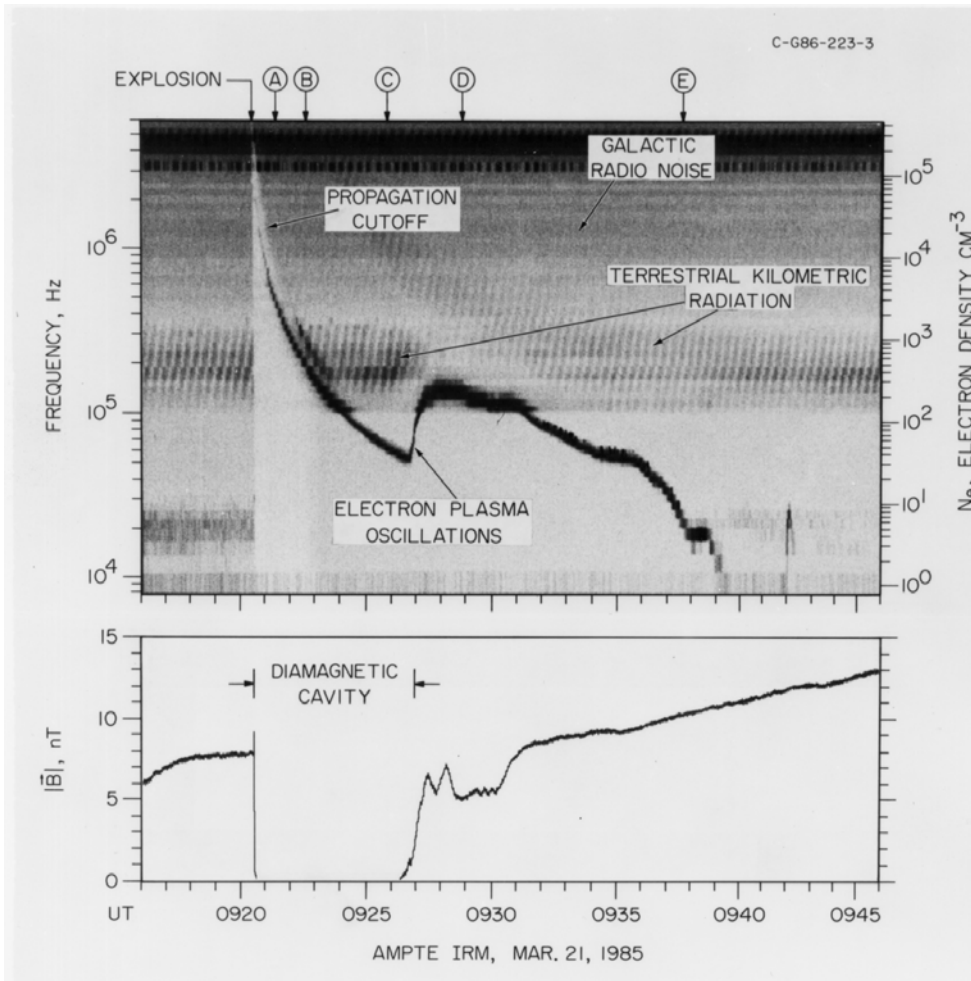


Figure 1.16 This image is from the paper: Observations and Theory of the AMPTE Magnetotail Barium Releases, P. A. Bernhardt et al, Journal of Geophysical Research, vol 92, page 5777, 1987. The figure was obtained from Dr. Don Gurnett, University of Iowa.

"Frozen-in" Magnetic Field Lines:

Consider a plasma which is embedded in a magnetic field. The field will in general be non-uniform as shown. If now either the plasma or the magnetic field configuration moves there will be an induced electromotive force generated within the plasma according to Faraday's Law

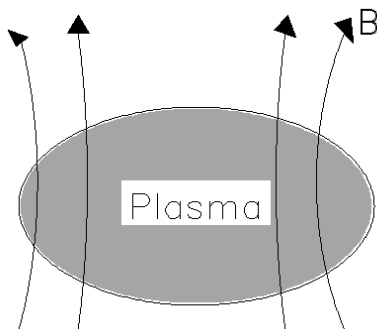


Figure 1.17

$$\mathcal{E} = -\frac{d\Phi_B}{dt} = -\int \frac{\partial \vec{B}}{\partial t} \cdot d\vec{A}$$

(Eqn. 1.42)

This electromotive force will generate an electric current within the plasma in such a direction that the magnetic field generated by the induced current opposes the change in magnetic flux which caused the electromotive force (Lenz's Law). If the conductivity of the plasma is high enough the induced currents and magnetic fields can become large enough to prevent any change in the external magnetic field. When these conditions apply we say that the field lines are "frozen" into the plasma and as the plasma moves the field lines will follow.

There exists a useful criterion which tells us when the "frozen-in" conditions are applicable. We define a quantity called the magnetic Reynolds number by

$$R_M = \mu_0 L v \sigma \quad \text{where} \quad (\text{Eqn. 1.43})$$

- L = Approximate size of the plasma system (m)
- v = velocity of fluid (m/sec)
- σ = electrical conductivity of fluid (Siemens/m = (ohm - m)⁻¹)
- μ_0 = $4\pi \times 10^{-7}$ (H/m)

R_M is a dimensionless quantity (pure number)

If $R_M \gg 1$ the B field will be frozen into the fluid and will move with it.

If $R_M \ll 1$ the fluid does not appreciably affect the B field.

We thus have two parameters to consider when discussing magneto plasmas:

The magnetic Reynolds number R_M tells us whether the field is frozen in the plasma or not. The parameter β tells us whether the motion of the combined plasma-field system is dominated by the plasma (large β) or by the magnetic field (low β).

E References

Physics of Fully Ionized Gases, Lyman Spitzer, Interscience Publishers, NY. , 1956.

Introduction to Plasma Physics, Francis Chen, Plenum Press, N. Y., 1984

Fundamentals of Atomic Physics, Atam Arya, Allyn and Bacon, Inc., Boston, 1971.

Atomic and Space Physics, Alex E Green and Phillip Wyatt, Addison-Wesley Pub. Co., Inc., Reading, MA, 1965.

Introduction to Modern Physics, F. K. Richtmyer, E. H. Kennard, John N Cooper, McGraw Hill Book Co., NY, 1969.

Physics of Space Plasmas, George K Parks, Addison-Wesley, Redwood City, CA, 1991.

Plasma Diagnostics, edited by W. Lochte-Holtgreven, Wiley & Sons, NY, 1968.

F Problems

1. What is the energy per photon for a 2 giga-Hertz (GH) signal, in eV? In Joules?
2. What is the frequency of a 1 mm wavelength electromagnetic wave?
3. For an (ideal) spherical satellite, of radius 1 meter, calculate the amount of power radiated into space for a vehicle operating at 20 °C. Remember you have to convert to degrees Kelvin.
4. At what wavelength does the radiation from the above satellite peak? If it was a re-entry vehicle at 1000 °C, at what wavelength would the radiation peak?
5. Calculate the equilibrium temperature of a satellite in free space, at a distance from the sun of 1 astronomical unit (1 AU = 93 million miles = 1.5×10^8 km, Solar Radius $R_{\odot} = 6.96 \times 10^5$ km.). Take the emissivity, ϵ , to be one for the sun (a good assumption), and the satellite (less so).

a) Calculate the radiated power: $P = \sigma T_{\text{sun}}^4 \cdot 4\pi R_{\text{sun}}^2$

b) Calculate the amount of power which reaches earth - it drops off as Gauss' law might suggest:

$$P = \sigma T_{\text{sun}}^4 \cdot 4\pi R_{\text{sun}}^2 \cdot \frac{1}{4\pi R_{\text{Earth Orbit}}^2}$$

You can check that this gives the known answer: $1370 \pm 4 \text{ W/m}^2$

c) The radiation incident on the satellite must be re-radiated:

$$P_{\text{incident}} = \sigma T_{\text{sun}}^4 \cdot 4\pi R_{\text{sun}}^2 \cdot \frac{1}{4\pi R_{\text{Earth Orbit}}^2} \cdot \pi R_{\text{satellite}}^2 = \sigma T_{\text{satellite}}^4 \cdot 4\pi R_{\text{satellite}}^2$$

noting that the radiation is incident on the projected area of the sphere, which is the circular disk cross-section. Solve the last equation for the satellite temperature.

6. For a 'Bohr' hydrogen atom, calculate the energy for the $n = 3$ to $n = 2$ transition. (Express the answer in eV). What frequency (and wavelength) does this correspond to? Is this wavelength visible?
7. The temperature of the gas in the upper atmosphere can reach 1000 K. Assume the region at this temperature is composed of hydrogen atoms. What is the average kinetic energy of one atom, in eV? What is the velocity of this atom (km/s)?
8. Evaluate the Maxwellian distribution, $n(K)$, for values of $K = 0.1 \text{ kT}$, 0.5 kT , 1.5 kT , and 10 kT . Take $N = 1 \times 10^6$, $kT = 1.0$ for convenience.
9. Calculate the Debye Length for a plasma with a temperature of 1,000 K, density = 1×10^5 electrons/cm³. (upper ionosphere, altitude about 1000 km). Calculate the Debye length at geosynchronous orbit: $T = 500 \text{ eV}$, $n = 1 \text{ electron/cm}^3$.
10. For a sphere of radius 0.1 m, with a charge of $1.11 \times 10^{-10} \text{ C}$ (111 pico-coulombs), plot the electric potential as a function of distance, from $r = 0.1 \text{ m}$, to $r = 2.0 \text{ m}$, in vacuum. On

the same plot, plot the potential if there is a plasma present, with density 2.0×10^8 electrons/m³, and a temperature of 1000 K. Repeat for a temperature of 1 eV.

Note that for an object of non-zero radius, the formula in the text needs to be recast into the form:

$$\Phi = \frac{q}{4\pi\epsilon_0 r} \exp\left(-\frac{r-r_0}{\lambda_D}\right).$$
 Here, the correction factor of $\exp\left(-\frac{r-r_0}{\lambda_D}\right)$ causes the potential at the surface of the sphere to be the same as it would have been, in the absence of a plasma.

11. Calculate the gyrofrequency (in Hz, or revolutions/second) for a proton in a magnetic field of 100 nano-Tesla (nT). (Note: a nT is also referred to as gamma, γ .)
12. Calculate the gyrofrequency (in Hz) for an electron near the earth's surface, near the equator. (You need to look up the earth's magnetic field....try chapter 4)
12. Calculate the drift velocity for a 5 eV electron, given an electric field of 1 milliVolt/meter (mV/m), which is perpendicular to a magnetic field of 200 nT. How does the drift velocity vary with energy?
13. Calculate the plasma pressure for the typical plasmasphere conditions: ion density = electron density = 2.0×10^{10} electrons/m³, ion temperature = electron temperature = 0.5 eV. Calculate the magnetic energy density for a field of 3 micro-Tesla ($3 \mu\text{T}$). Calculate beta (β).

This page intentionally left blank

Chapter 2 The Sun

A Introduction

The sun is a modest sized star, located at the edge of the milky way galaxy. It is important to us as a source of energy. More particularly, the sun is a source of light and radiation (photons), and charged particles. We need to understand the normal behavior of the "quiet" sun, and the nature of the variations in the sun's behavior. The radiative output of the sun determines the behavior of the earth's neutral atmosphere. A more subtle, but important effect of the UV radiation from the sun is the formation of the ionosphere. Variations in the sun's UV output affect the ionosphere, and hence communications. The particle output from the sun (the solar wind) results in the region of trapped plasmas around the earth, the magnetosphere. Variations in the solar wind result in magnetic storms, which dramatically affect satellites at geosynchronous orbit.

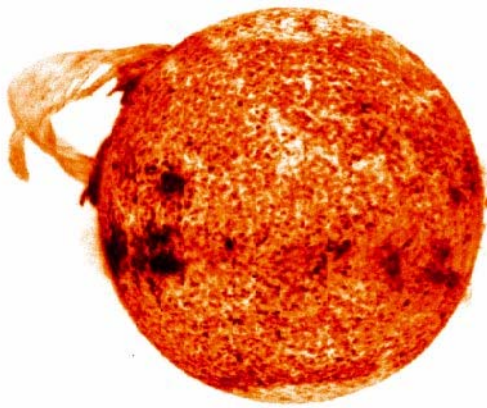


Figure 2.1 This photograph of the sun was taken using the extreme ultraviolet radiation from ionized helium, at 304 Angstrom. It was taken on December 19, 1973 using the NRL Solar Physics Branch's SO82A Spectroheliograph onboard the Skylab ATM. Note that the image here is a negative, to enhance visibility - the small black regions are bright in He II.

B A few facts about the quiet sun

Solar Radius $R_{\odot} = 6.96 \times 10^5 \text{ km} = 109 R_{\oplus}$ which means that the volume of the sun is about 1.3 million earth volumes. (R_{\oplus} is the radius of the earth)

Distance from earth to sun: Average $1.5 \times 10^8 \text{ km}$ which is $215 R_{\odot}$. Earth moves in elliptic orbit with perihelion distance of $1.47 \times 10^8 \text{ km}$ and aphelion distance of $1.52 \times 10^8 \text{ km}$. Period = 365.256 earth days.

1 Astronomical Unit = 1 AU = $1.49598 \times 10^8 \text{ km}$ ~ the distance from the earth to the sun

Solar Mass $M_{\odot} = 1.989 \times 10^{30} \text{ kg}$ giving an average density of $1.4 \times 10^3 \text{ kg/m}^3$ or 1.4 times density of water at ordinary temp. and pressure.

Composition: 70% hydrogen, 28% helium, 2% other (by mass). Total of 62 elements have been identified.

Overall magnetic field of 100 - 300 μT . Localized B fields can be up to 100 times these values.

Estimated Age of Sun ~ 5×10^9 years, Life expectancy 10 - 15×10^9 yrs. (If nothing unexpected comes up)

Electromagnetic Power radiated $3.86 \times 10^{26} \text{ W}$ (total). This gives for power/ m^2 at 1 AU: $1370 \pm 4 \text{ W/m}^2$. This quantity is called the solar constant.

Spectral Distribution: 10⁻⁸% RF, 52% IR, 41% Visible, 7% Near UV, 0.1% Far UV and X-Rays.

Solar Wind Emission: 10^9 kg/sec total from solar surface, mostly protons & electrons.

We have already encountered the solar spectrum in chapter 1, in Figure 1.7. There it was shown over a wide range of wavelengths (from the near UV to the far IR) the sun's spectrum as observed in space closely agrees with a blackbody at about 5800 K. The peak of the distribution curve lies near the center of the visible wavelength region. The "effective" temperature, as defined by the Stefan-Boltzmann law ($R = \sigma T^4$), is 5800 K. The shape of the curve is better described by a black-body curve for an object at 6000 K. The discrepancy is due to variations in the temperature within the photosphere, and limb effects. (See Kenneth Phillips, *Guide to the Sun*, Cambridge Press, 1992, pages 83-84; Cambridge Encyclopedia of Astronomy, pages 131-132).

Superimposed on this black-body radiation curve it is found that there are narrow absorption bands termed Fraunhofer lines, which result primarily from absorption in the solar atmosphere. From them, we can diagnose the composition and temperature profile of the solar atmosphere. In addition, there are a wide variety of effects in the solar atmosphere which produce non-thermal radiation signatures.

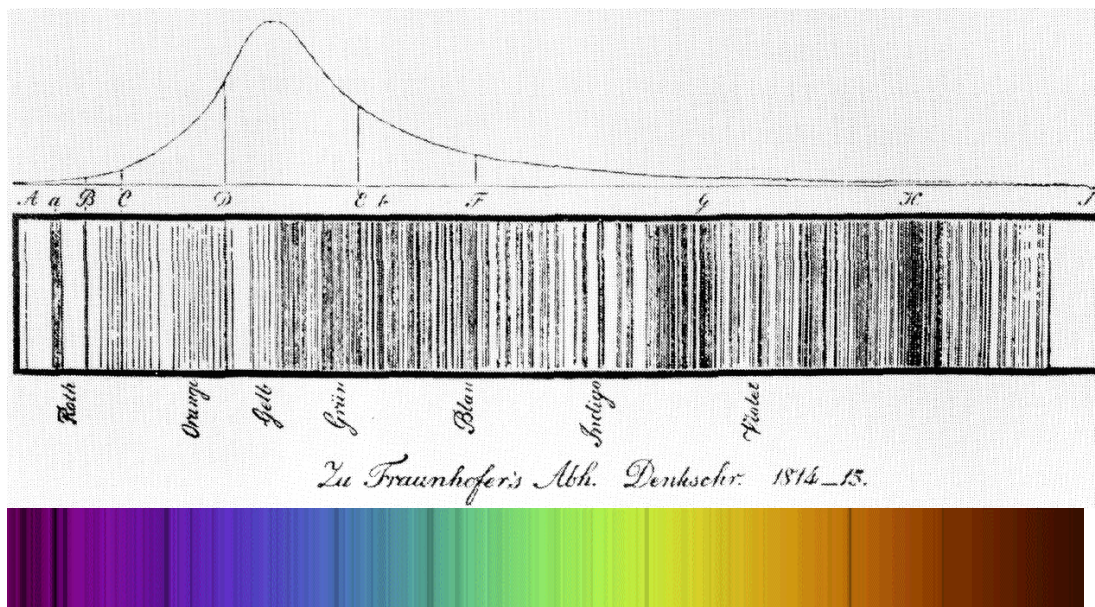


Figure 2.2 Fraunhofer spectrum of the sun. Top figure from: Phillips, *Guide to the Sun* A similar figure is in *Sun and Earth*, Friedman, page 15, 1986, and *Introduction to Physics*, Jay Pasachoff and Marc Kutner, WW Norton and Company, NY, 1981, plate 17. The figure in Pasachoff and Kutner is estimated to come from 1814. Bottom figure: Institut National des Sciences de l'Univers / Observatoire de Paris; BASS 2000 - BAsse Solaire Sol 2000 - Antenne meudonnaise; http://mesola.obspm.fr/form_spectre.html

C The Structure of the Sun

As can be seen in Figure 2.3 we can divide the interior of the sun into several concentric regions and discuss each region in terms of its contribution to the energy transport from the core to the "surface" and beyond. The core of the sun extends to $0.2 R_{\odot}$ and is the region in which fusion occurs. Hydrogen is converted into helium in the core, according to processes described in a later section. Surrounding the core is a rather massive region extending to approx. $0.8 R_{\odot}$ in which radiative transport dominates. In this radiative transport region ($0.2 R_{\odot} < r < 0.8 R_{\odot}$) there is relatively little change in temperature and hence not much mass motion of the plasma. The energy leaving the core is mostly in the form of high energy ($\sim \text{MeV}$) photons. As these photons work their way outward they are absorbed and reemitted many times, each time at a somewhat lower energy. This process is sometimes referred to as radiative diffusion and consists of a large number of discrete steps in a "random walk" pattern. As a result of these multiple interactions the average photon energy is gradually decreased from γ rays to x rays to ultraviolet and eventually down to the visible light range.

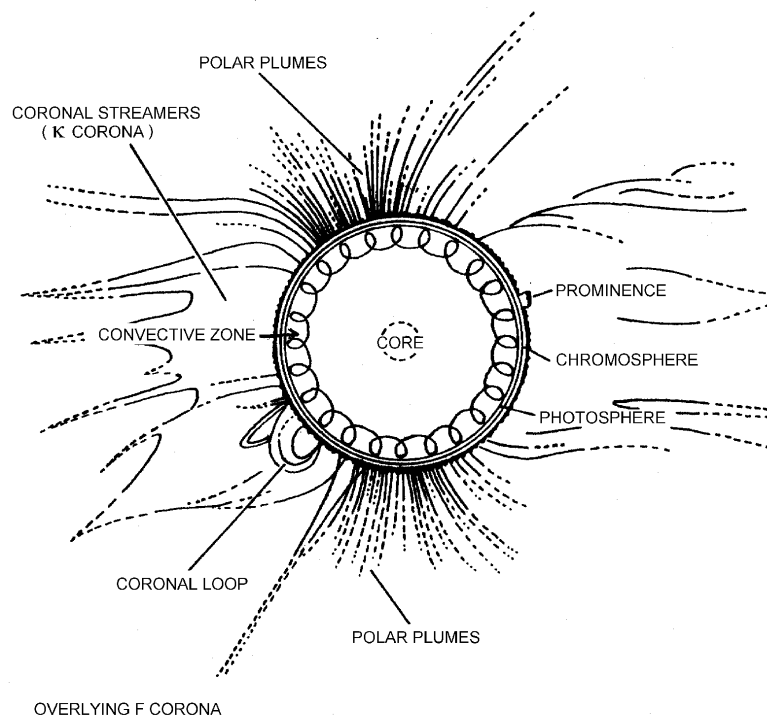


Figure 2.3: STRUCTURE OF THE SUN and the corona is summarized in this cross-sectional diagram. Energy from thermonuclear reactions in the core makes its way gradually to the sun's exterior layers. Most of the visible light received on the earth comes from the photosphere. Just under the photosphere is the convective zone. Shock waves from the convective zone carry energy up into the chromosphere and the corona. The temperature of the sun is at a minimum in the photosphere and lower chromosphere; the shock waves cause the temperature to rise through the upper chromosphere and the corona until it reaches some two million degrees Kelvin in the corona. From the corona the solar wind expands into interplanetary space. The visible corona has three major components. The first is the K corona: light scattered by the electrons in the gas surrounding the sun. The second is the F corona: light scattered by the interplanetary dust between the sun and the earth. The radiation of the third component is emitted by highly ionized atoms near the sun.

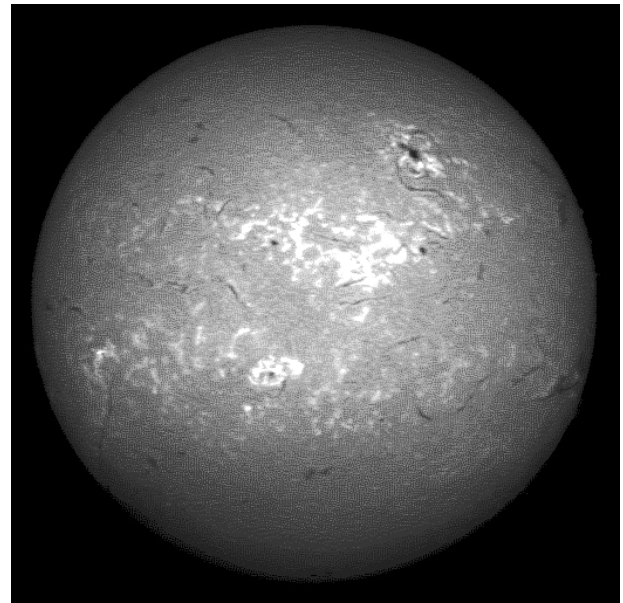
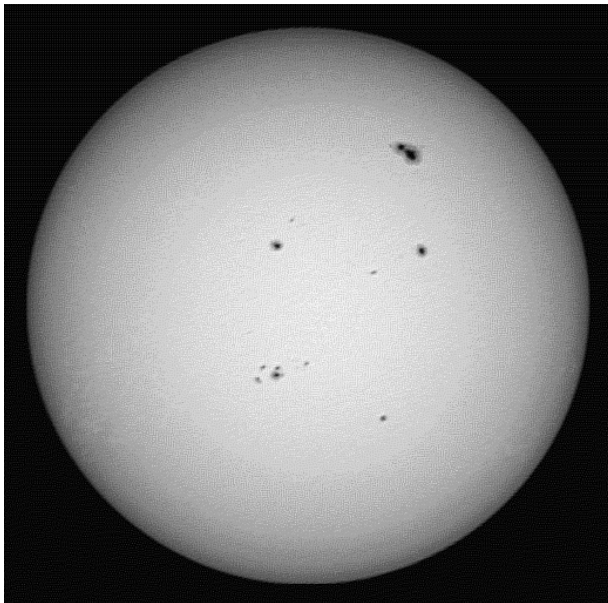
From: Scientific American, October 1973, p 69, The Solar Corona, by Jay N. Pasachoff

In the outer 'convective' region, transport of energy is via motion of the plasma, in a turbulent process which still shows a surprisingly well defined structure, defined by granular cells which are visible at the visible surface of the sun, termed the photosphere. At the surface, a great deal of structure can be seen, with features such as sunspots, plages, pores, and granulation. Immediately above the photosphere is region of cooler plasma, extending upwards some 10,000 km, termed the chromosphere. The solar temperature reaches a minimum of about 4,700 K at the lower boundary of this region.

Extending outward from the surface into the chromosphere (and beyond) are prominences and flares, the active processes associated with magnetic activity. Above the chromosphere is the corona, normally only visible during a total eclipse of the sun. Here the solar atmosphere increases in temperature to 1-2 million degrees.

The highly luminous surface of the sun is called the "photosphere". The photosphere is the sharp disc as observed with the eye or a small telescope. In white light, the edge of the sun, or "limb" appears as a sharp boundary. This is typically taken as giving the diameter of the sun, 1.391×10^6 km. Practically all of the mass of the sun is contained within this boundary. The white light image below (Figure 2.4a) shows the light emitted from that surface.

The photosphere is only ~ 500 km thick. Below the photosphere, the gas is opaque; above the photosphere, the gas is transparent. Most of the emission is from H^- (hydrogen with an extra electron). The H^- emission provides the continuum characterized by us as white light, as imaged in Figure 2.4a. (It is not an accident that there is such a sharp edge – the gradual drop in temperature and density reaches a point where radiation can escape – that escape then allows for a further very rapid cooling of the plasmas, allowing for the increasingly effective escape of the emitted radiation.)



a) White Light - photosphere

b) $H\alpha$ (6563 Å) - Chromosphere

Figure 2.4 Solar images from Big Bear Solar Observatory (BBSO), acquired on 13 May 1991.

Measurements over more limited wavelength ranges, at specific wavelengths, allow us to image the sun at different altitudes above the visible surface. Figure 2.4b shows one such observation, the so-called “H-alpha” measurement at 656.3 nanometer (6563 Angstroms or Å). This is one of the Balmer series transitions illustrated in Figures 1.2 and 1.5. The image primarily shows the effects of absorption by neutral hydrogen above the photosphere, a process which is largely localized in regions which are 1200-1800 km above the visible surface of the sun (e.g. in the chromosphere). (The H α transitions are associated with temperatures of about 6000 K.) Above active regions on the Sun’s surface (such as those found around sunspots), the chromosphere is heated, and the H-alpha intensity is higher (plage regions). Narrow dark ‘filaments’ are also observed, as talked about below.

When we observe the surface of the sun at high magnification, it reveals a mottled texture termed “granulation”. Gradations in this structure exist, but the smallest granules consist of bright patches of light about 1000 km across, with a dark border. These granules have a lifetime of 5 - 15 minutes, and represent an important indication of the turbulent convective motion which is present just below the visible surface of the sun. Figure 2.5 shows an image of this fine scale structure. Doppler shift measurements show that the bright center of the cells is moving upwards. The center regions are typically 100 - 500 K hotter than the edges, where the gas descends. Vertical gas velocities are on the order of 2 - 3 km/s. They are imbedded in larger scale structures which have been dubbed ‘super-granulation’ cells. These have a characteristic size of 30,000 km, and a nominal lifetime of 20 hours. These granulation cells will prove important as an element in the formation of the solar wind, as described in the next chapter.

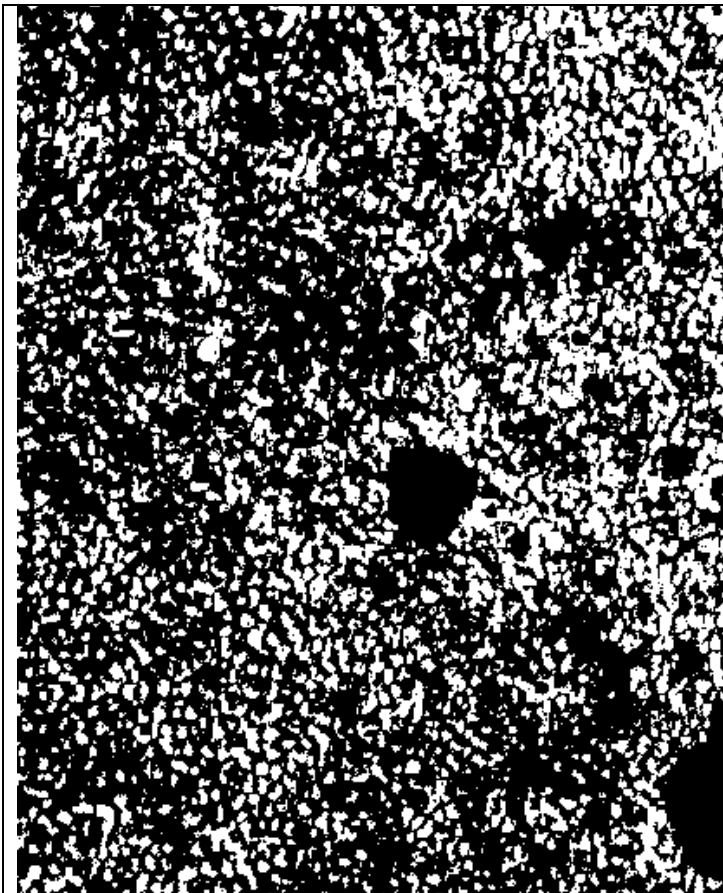


Figure 2.5a Solar granulation, pores, and small spots, July 5, 1885 (Janssen, 1896), 1mm = 0.5". (This may be the best photograph of solar granulation in existence. Janssen wrote about it: "obtenue sans aucune intervention de la main humaine.") Taken from: The Sun, edited by Gerard P. Kuiper, 1953, Chapter 6, Solar Activity, by K. O. Kiepenheuer, p 341.

SOHO MDI/SOI
1996 May 26
00:00 UT

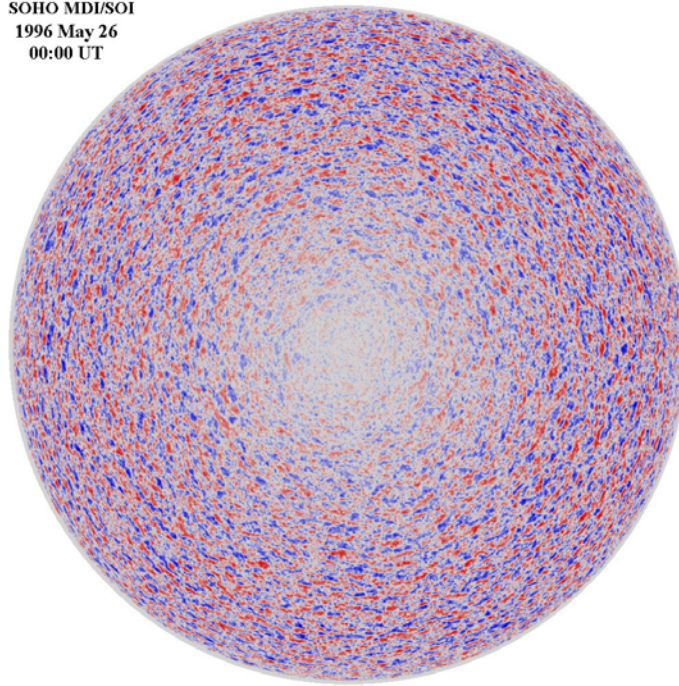


Figure 2.5 b

Supergranules are much larger versions of granules (about 35,000 km across). They are best seen in measurements of the "Doppler shift" where light from material moving toward us is shifted to the blue while light from material moving away from us is shifted to the red. These features also cover the entire Sun and the pattern is continually evolving. Individual supergranules last for a day or two and have flow speeds of about 0.5 km/s (1000 mph). The fluid flows observed in supergranules carry magnetic field bundles to the edges of the cells where they produce the chromospheric network.

<http://science.nasa.gov/ssl/pad/solar/feature1.htm>

Figure 2.6 shows some of the consequences of the granulation structure. Note that the spicules (the 'burning prairie' of the chromosphere) occur at the boundaries 'super-granulation' cells. Note that spicules have a characteristic size of about 1000 km across, and have a lifetime of about 4 minutes. Very recent work (1999) indicates that there is an important concentration of magnetic flux in such regions, and significant upward plasma flow at the boundaries of the chromospheric network.

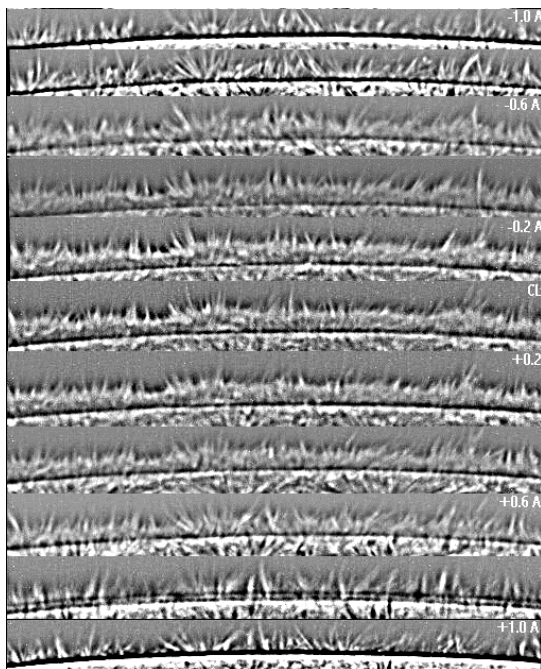


Figure 2.6a. (top) The limb of the Sun at different wavelengths within the Hydrogen-alpha spectral line are shown. Images taken from 1 Angstrom blue-ward to 1 Angstrom red-ward of the line center. Some of the spicules (jets) extend above a height of 7000 km. The images have been processed with a high pass filter. Image from Big Bear Solar Observatory

<http://sundog.caltech.edu/daily/image.html>.

See also, Our Sun, Menzel, page 164.

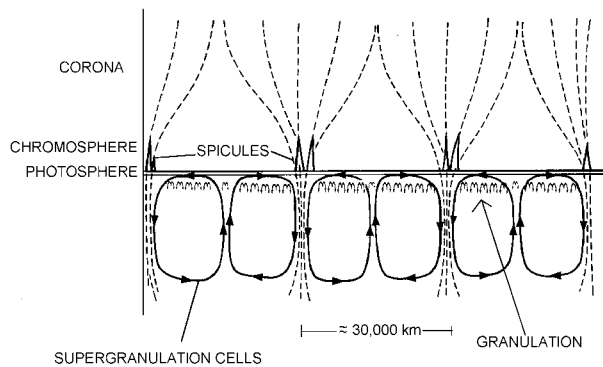


Figure 2.6b Structure at the solar surface. The flow of ionized gas from the center to edge in (super-)granulation cells carries magnetic fields to the cell boundaries, and a network of enhanced field results. Spicule location coincides with this network. (Einar Tandberg-Hansen, figure 3.2, 1967). This larger scale (granulation) structure is visible in images at 3933 Å (Ca II K image), and is termed the chromospheric network. The upward/downward motion is visible in 'spectro-heliograms', or Doppler images. The Sun, Our Star, Robert Noyes, pages 134-138.

D Solar Atmosphere

Below the photosphere the solar gas is opaque. This opacity is primarily due to a small concentration of negative hydrogen ions in the region immediately below the photosphere. These negative ions act as continuous absorbers over a great range of wavelengths, absorbing most of the intense radiation from deeper layers in the sun. The energy is reradiated into the relatively transparent gases above the photosphere and hence into space. Because the opaque gas composing the photosphere absorbs and reradiates approximately as a black body, the photosphere emits an essentially continuous spectrum. Superimposed on this continuum are a large number of dark or absorption lines (Fraunhofer lines) due to atoms of heavier elements both in and above the photosphere. This absorption spectrum is the source of most information on the abundance of the various elements in the sun and the physical state of the solar atmosphere. All the natural chemical elements known on earth probably exist in the sun, but with radically different relative abundances. Figure 2.7 shows the variation of temperature in a standard model of the solar atmosphere.

As we proceed outward from the base of the photosphere the temperature drops to a minimum of about 4700 K at an altitude of 500 km (see Figure 2.7). However beyond that point it begins to rise again as we enter the chromosphere which is a layer several thousand kilometers thick consisting of transparent glowing gas. The chromosphere is a thin, pinkish colored region when observed during eclipses. The solar atmosphere at these levels is diagnosed by observations of Ca II (3934 Å, 10,000 K) and He I at 10830 Å (50-200,000 K), as shown in Figure 2.8.

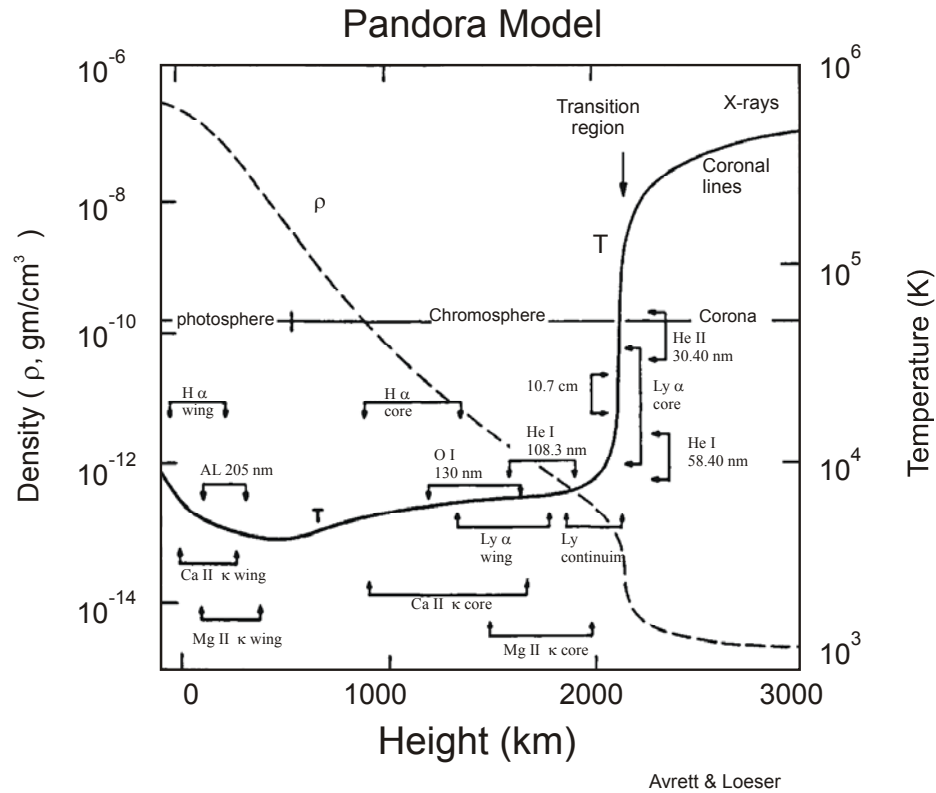


Figure 2.7 Temperature and density in the solar atmosphere versus height. The calculated position of the formation of several Fraunhofer lines are also shown. (From Avrett E.H. & Loeser R., 1992, 7 Cambridge Workshop on Cool Stars, Stellar Systems and the Sun, ASP Conference Series, 26, Eds. Giampapa M. & Bookbinder J.

1 Chromosphere

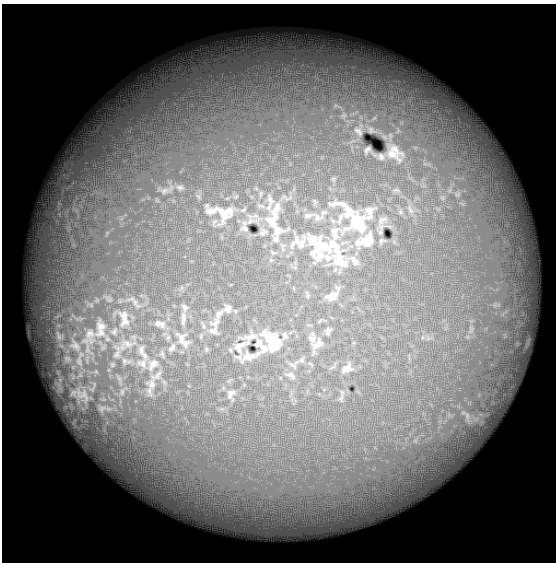


Figure 2.8a. Ca II; 3934 Å; 10,000 K

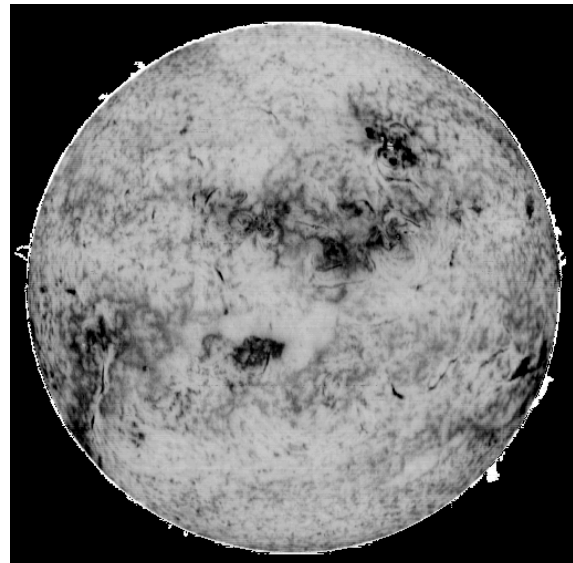


Figure 2.8b He I; 10830 Å; 50-200,000 K

The Ca II image is from Big Bear Solar Observatory (BBSO). The He I image is from NSO/Kitt Peak. Images acquired on 13 May 1991. Compare to Figure 2.4.

The chromosphere is illustrated here in a pair of images – one from the deep blue line at 3934 Å, the other an infrared image of light which is absorbed by a neutral helium region. The Ca II K line is a strong spectral line associated with once-ionized Calcium. It has a wavelength of 393.4 nm (billionths of a meter, in the blue part of the spectrum) and absorbs about 98 percent of the light at its central wavelength. In this spectral line you view layers of up to 2000 km above the visible surface of the Sun. The center of this spectral line is very sensitive to the presence of magnetic fields in the material. If a magnetic field is present, then the absorption is less, i.e. more light is transmitted. Moderately strong magnetic fields show up bright in images taken in this spectral line, but strong magnetic fields (such as in sunspots) don't. Typical Ca K images show brightness along the edges of cells (called supergranules) and in certain isolated areas (called plage) which, when enough magnetic field is present, are associated with sunspots and are then called active regions.

Prominences are readily seen at the limb. The bright ring at the limb is the portion of the chromosphere brighter than the sky intensity threshold. This figure basically reflects regions where the excited helium atoms absorb radiation from below. For contrast, the SOHO satellite routinely images the sun at 304 Å, showing radiated light from He II.

The chromospheric network is a web-like pattern most easily seen in the emissions of the red line of hydrogen (H-alpha) and the ultraviolet line of calcium (Ca II K - from calcium atoms with one electron removed). The network outlines the supergranule cells and is due to the presence of bundles of magnetic field lines that are concentrated there by the fluid motions in the supergranules. (<http://science.nasa.gov/ssl/pad/solar/feature2.htm#Network>)

2 Transition Region

At increasing altitudes, one observes the transition region, then the corona. The transition region is illustrated in Figure 2.8b above, and Figure 2.9, an image taken by SOHO at 977.02 Å, which is a Carbon III line. The transition region is a thin (a few hundred km) and very irregular layer of the Sun's atmosphere that separates the hot corona from the much cooler chromosphere. Heat flows down from the corona into the chromosphere and in the process produces this thin region where the temperature changes rapidly from 1,000,000°C (1,800,000°F) down to about 20,000°C (40,000°F). Hydrogen is ionized (stripped of its electron) at these temperatures and is therefore difficult to see. Instead of hydrogen, the light emitted by the transition region is dominated by such ions as C IV, O IV, and Si IV (carbon, oxygen, and silicon each with three electrons stripped off). These ions emit light in the ultraviolet region of the solar spectrum that is only accessible from space.

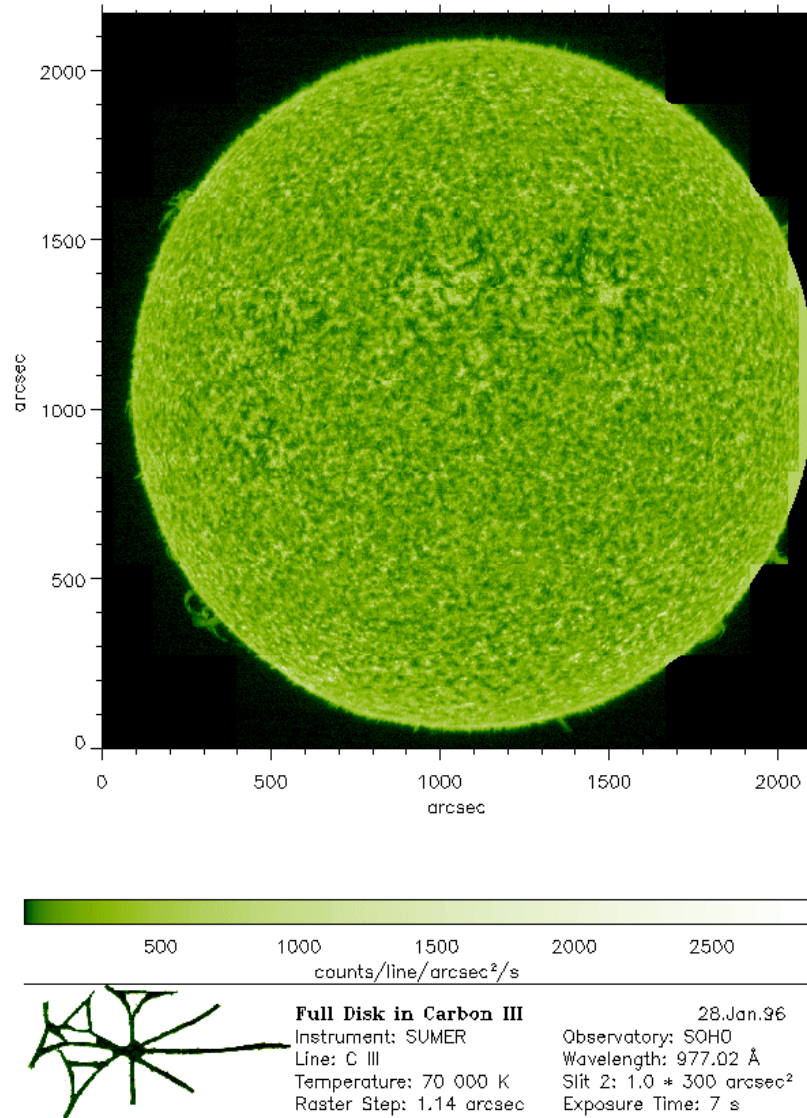


Figure 2.9 The Sun observed by SUMER on 2 February 1996 in the emission line of C III lines at 977.020, formed in the transition region at a temperature of about 70 000 K. The image is shown in bins of 4x4 pixels, one pixel being approx. 1 arcsec. The patchy pattern is the chromospheric network, with individual cells being of the size of about 30 000 km. Also note some prominences over the limb. This image was the first full Sun scan of SUMER.

<http://sohowww.nascom.nasa.gov/>

3 The Corona

At the interface between the atmosphere and the corona (which is the outer "atmosphere" of the sun) the temperature shoots up to $1 - 2 \times 10^6$ K and remains at this value throughout the corona. Figure 2.10 shows an image in Lyman- α . Extreme ultraviolet (EUV) and x-ray images such as these reveal the heating that occurs above magnetically active regions. Several mechanisms (Acoustic Shock Waves, Hydrodynamic Waves, etc.) have been proposed to explain this unexpected increase in temperature, but none of these theories is completely satisfactory at this time.

The high temperature of the coronal plasma (~ 100 eV) would normally produce an extremely bright region. However, because of the low density of the coronal plasma ($\leq 10^5$ particles/cm³), it is largely transparent and is only visible in white light during a total eclipse of the sun. The visible light is primarily continuum radiation from the sun, scattered by coronal electrons. Emissions from the corona itself are mainly in the x-ray and radio bands. The corona extends out several solar radii. Beyond this point, further evaporation of the sun's atmosphere produces the solar wind. Figure 2.10 illustrates the solar corona as seen during an eclipse from earth.

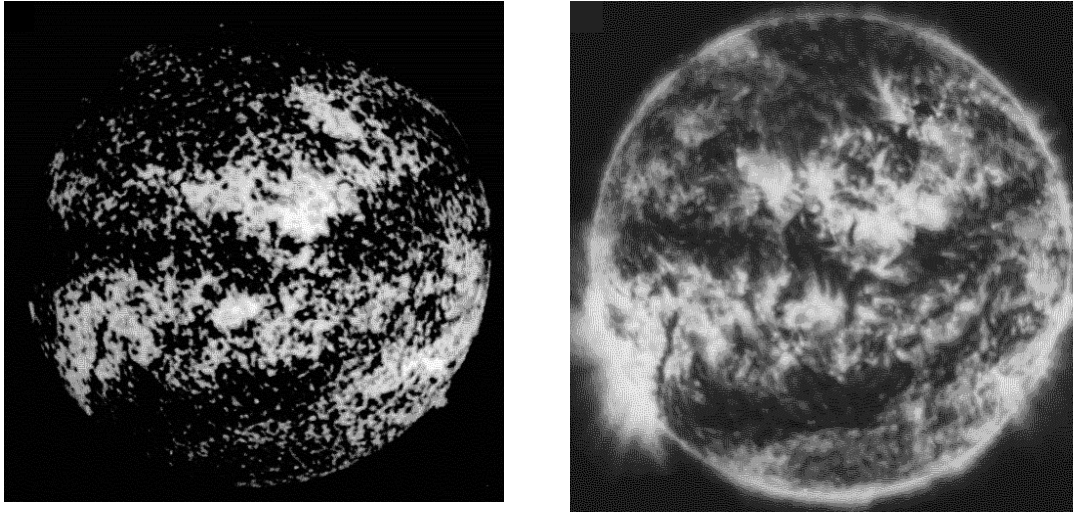


Figure 2.10a. MSSTA image in H-Lyman- α ; 1216 Å; 20,000 K; Figure 2.10b. MSSTA image in Fe XIV; 211 Å; 1,800,000 K; 13 May 1991

A set of high quality x-ray images was acquired by Richard Hoover (NASA/MSFC) and Art Walker (Stanford) on 13 May 1991, with the Multi-Spectral Solar Telescope Array (MSSTA). (Hoover, R. B., A. B. C. Walker, J. Lindbloom, M Allen, R. O'Neal, C. DeForest, T. W. Barbee, Solar Observations with the Multi-spectral Solar Telescope Array, SPIE Proceedings Volume 1546, Multilayer and Grazing Incidence X-ray/EUV Optics, page 175, 1991.)

The Lyman- α image from that experiment is shown here.

The upper atmosphere of the sun can be observed at other times (that is, not during eclipse) by specially equipped ground observatories, rocket experiments, and satellites. This is generally done in the extreme ultraviolet and in x-rays. An x-ray image associated with emission from excited iron ions is shown here from the MSSTA experiment. Note that Fe XIV is iron ionized thirteen times - the source of the (relatively) famous 5303 Angstrom coronal 'green' line that so puzzled solar astronomers - they could not determine what the element was, and for a while ascribed the spectral measurements (seen during solar eclipses) to the element 'coronium' (discovered by Young and Harkness during an eclipse in 1869; Phillips, Guide to the Sun, page 23).

(The eclipse of 1869 produced several coronal lines in the visible region, which could not be attributed to any element known on earth. They were attributed to a new element, 'coronium'. The brightest lines were the green line (530.3nm) and the red line (637.4nm), with a few dozen fainter lines also discovered. The actual source of these was finally explained in 1939 as due to forbidden transitions in highly ionized iron (Fe XIV and Fe X respectively). The high temperatures necessary to create these ionization states, and the low densities to ensure the forbidden transition from the metastable states, provided further evidence of the extraordinary conditions in the corona.)

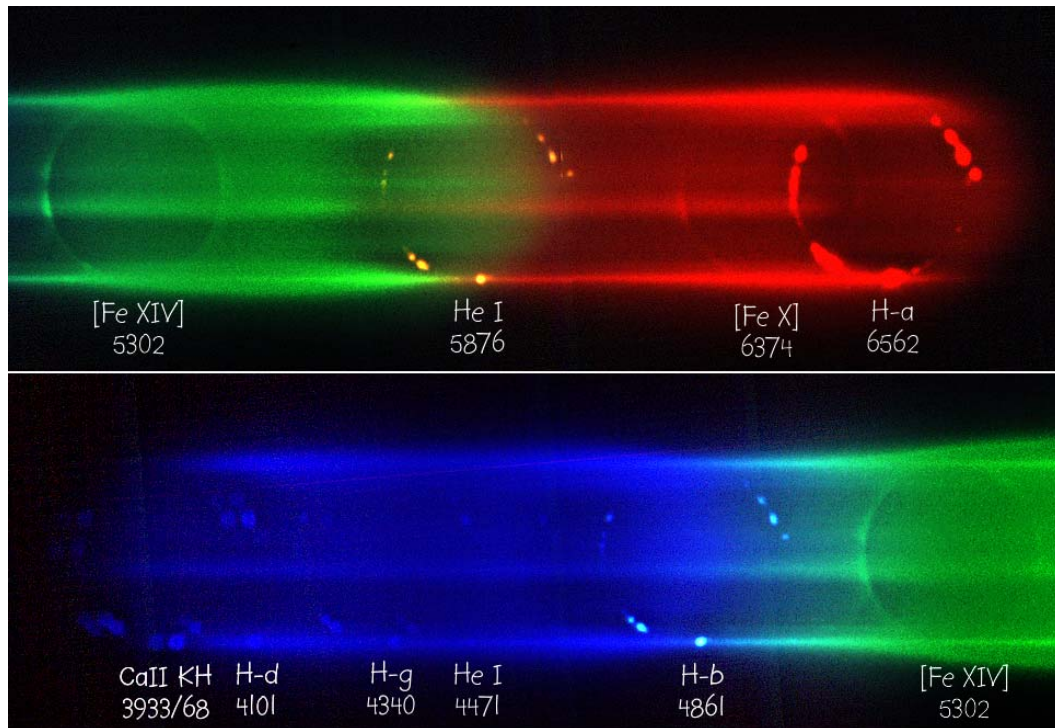



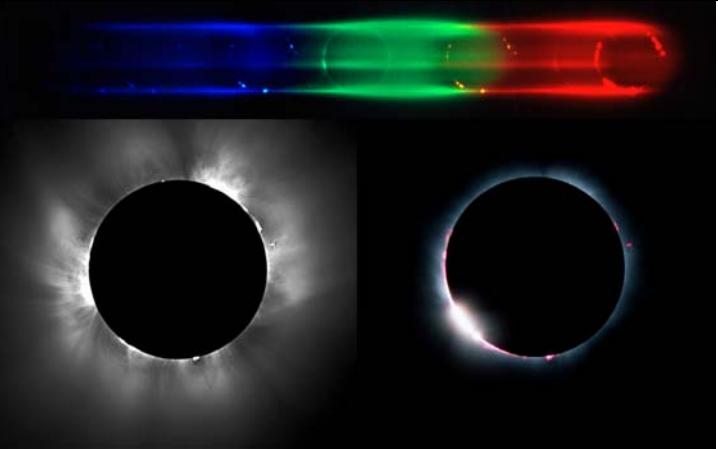
Figure 2.11

The dominant emission lines are the Balmer series from H-alpha to H-delta and the Helium D3 line at 5876A. The two strong coronal lines [FeXIV] 5303 and [FeX] 6374 can be seen with a different spatial distribution. Two images show – one of the red (1/2 sec) and blue (1 sec) exposures taken approximately in mid-eclipse. These are labeled with the hydrogen, helium, iron and calcium lines which are clearly visible. The calcium H and K lines are in a region where the lens focus is not perfect. Note the high prominence to the west (right) seen clearly in H α and D3. The chromospheric (hydrogen, helium, calcium) and coronal (iron) emission lines have quite different spatial distributions.

<http://ecf.hq.eso.org/~rfosbury/home/photography/Eclipse99/csp.html>



Spectra shown with a direct image taken at 3rd contact by Philippe Duhoux from a site NW of Munich (right side of image). The prominences and the bright low coronal regions can be easily identified. CCD coronal image (left) taken in France at Vouzier (Champagne-Ardennes) by Cyril Cavadore from ESO and L. Bernasconi and B. Gaillard from Obs. de la Cote d'Azur.



http://ecf.hq.eso.org/~rfosbury/home/photography/Eclipse99/csp_cor_chr.jpg
 Direct images from ESO Report about the Solar Eclipse on August 11, 1999
<http://www.eso.org/outreach/info-events/eclipse99/report-hq.html>



Figure 2.12 This image of the Sun's corona was made from a composite of eight separate photographs made by Fred Espenak from Dundlod, India during the total solar eclipse of 1995 October 24. The photos were made on Kodak Royal Gold 100 with a Nikon FE w/MD-12 motor drive, a Sigma 400mm f/5.6 APO telephoto and a Sigma 2X teleconverter. Exposures were 2, 1, 1/2, 1/4, 1/8, and 1/15 and 1/125 seconds. The eight images were combined into one composite image in order to show the corona properly. Photo ©1996 Wendy Carlos and Fred Espenak.

http://umbra.nascom.nasa.gov/eclipse/images/eclipse_images.html

The best way to see the corona, however is in x-rays. This has been done with film, from skylab, and via electronic means, as with Yohkoh.

Figure 2.13 The corona X-Ray Spectrographic Telescope (S054) on Skylab provided some of the earliest glimpses of the corona, including the remarkable "boot of Italy" coronal hole structure which persisted for several solar rotations. Image is formed at 2-32 and 44-54 Å With thanks to the efforts of Dave Batchelor at keeping the data set alive.

http://nssdc.gsfc.nasa.gov/nssdc_news/march95/06_d_batchelor_0395.html

Further information can be found in A New Sun: The Solar Results From Skylab by John A. Eddy (Publ. by National Aeronautics and Space Administration, Washington D.C., 1979).

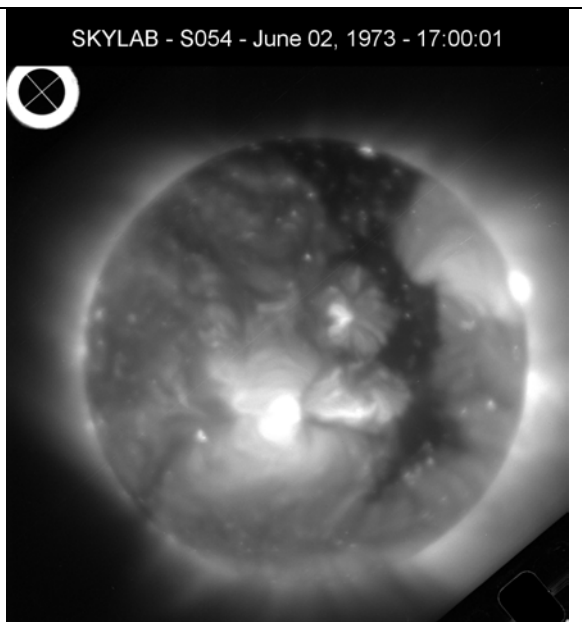
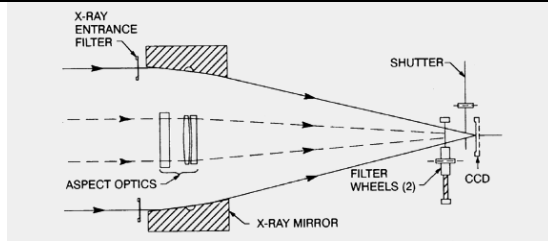
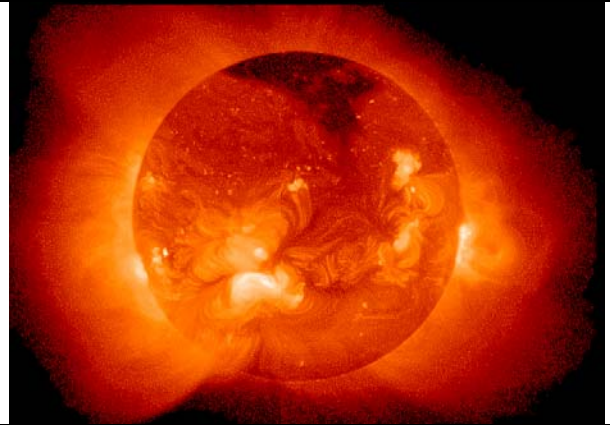


Figure 2.14 This is an X-ray image of the sun taken on 26 August 1992. The image was made up from two pointings of the spacecraft, one to the east and one to the west, to capture the distant corona far above the sun's limb. Again, a coronal hole is obvious at the northern solar pole.
(X-ray Filter : AlMg ?)

Compare the energy ranges for Yohkoh and Skylab



The soft X-ray telescope (SXT) is a glancing incidence telescope of 1.54 m focal length which forms X-ray images in the 0.25 to 4.0 keV range on a 1024x1024 virtual phase CCD detector. A selection of thin metallic filters located near the focal plane provides the capability to separate the different X-ray energies for plasma temperature diagnostics.

<http://www.lmsal.com/SXT/homepage.html>

E Solar Magnetism

1 Introduction

The general magnetic field of the sun has intensity of about $100 \mu\text{T}$ (~ 1 Gauss) at the surface, which is slightly higher than the intensity of the earth's magnetic field ($30 - 60 \mu\text{T}$). The magnetic field of the sun is largely a surface phenomenon, and the sun's magnetic field should not be thought of in the bar magnet analogy which is so useful for the earth. There is an overall polarity to the sun's field, but it is not steady in time. The most obvious manifestation of the sun's magnetism is in sunspots.

2 Sunspots

Figure 2-15 shows a modern image. Sunspots look like irregular holes in the sun's surface. There is a black inner region, the umbra, with a more luminous fringe, the penumbra. A typical spot is about 10,000 km across, but they have been observed up to 150,000 km. They often occur in pairs (about 90% of the time), and these pairs appear in groups. Sunspots are photospheric regions where the magnetic flux is concentrated and field strengths are on the order of 0.2 to 0.4 Tesla (4000 Gauss). In this region of intense magnetic fields, the temperature, radiation and gas pressure are reduced. The reduced temperature (~ 3900 K) causes the spot to appear darker than the surrounding hotter photospheric gases. (The spot is actually about as bright as the full moon. Observations of sunspots as they approach the solar limb reveals that they are depressions in the surface of the sun.

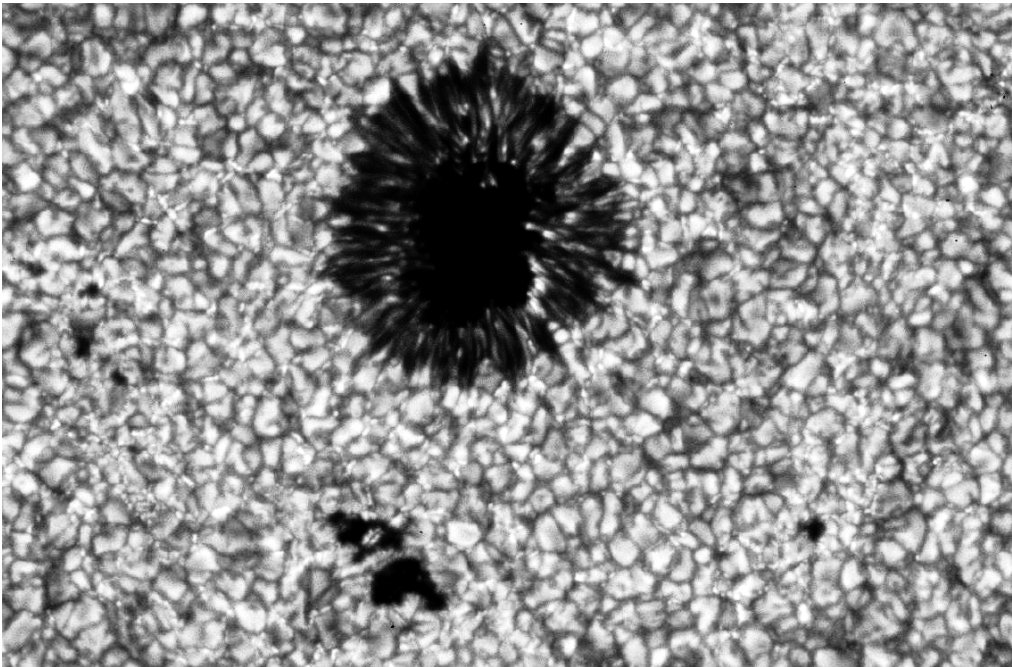


Figure 2.15 White light (continuum) image, 14 June 1994. One sunspot, with a background of granulation. Pores are visible near the bottom of the image. Source: Kiepenheuer/Uppsala/Lockheed (P. Brandt, G. Simon, G. Scharmer, D. Shine) (Found in WWW).

The magnetic nature of sunspots is revealed in 'magnetograms', measurements of the solar magnetic field obtained from spectral measurements of (polarized) emissions showing the Zeeman Effect - a spectral feature which arises because of the interaction of magnetic fields with the magnetic moments of the electrons. (An excellent discussion of solar magnetism, and the measurement thereof, is given in Our Sun, Menzel, pages 108-117)

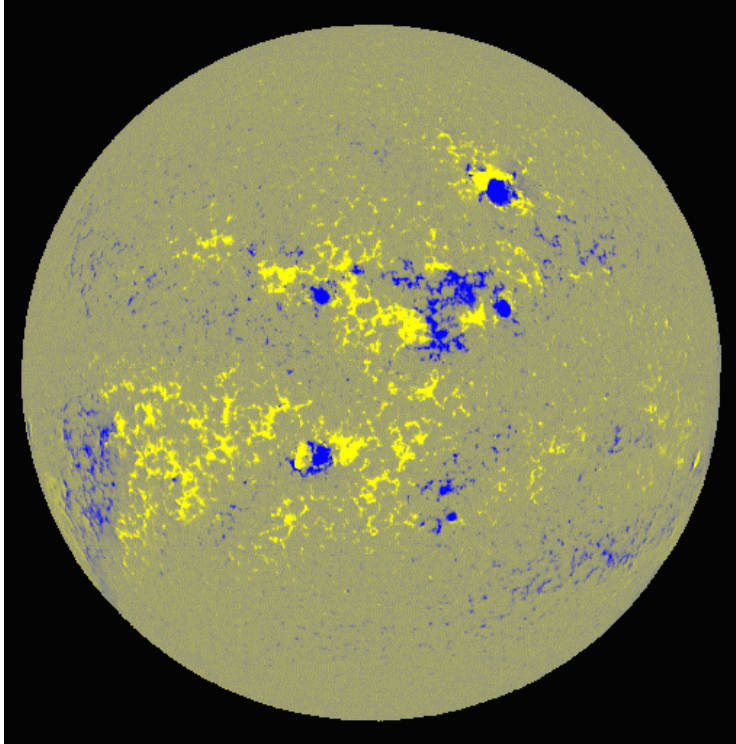


Figure 2.16 Magnetic field measurements, 13 May 1991.

Upon comparing this magnetogram to the $H\alpha$ or white light images above, one sees that regions of strongest magnetic fields (yellow and light blue on the magnetogram) always coincide with sunspots. Diffuse magnetic fields of lesser strengths are also present all over the solar surface, with moderately strong (Gauss) fields most often associated with plages. Field strengths in sunspots are in the range 1000--4000 Gauss, with the stronger fields in the larger sunspots; this is much larger than the average 0.5 Gauss of the Earth's surface magnetic field. The magnetically active regions also correspond to the hotter regions revealed in the x-ray observations.

Sunspots are almost never seen in complete isolation, but instead are most often grouped in pairs of opposite magnetic polarities. Isolated sunspot pairs tend to line up in the East-West direction (approximately from left to right on this magnetogram). Further scrutiny of magnetograms such as this one reveals that the magnetic polarities of sunspot pairs located in the northern and southern solar hemispheres are reversed; in one hemisphere the negative magnetic polarity sunspot almost always leads the positive polarity sunspot (with respect to the westward apparent motion due to solar rotation), while a similar behavior, except for reversed magnetic polarities, is observed in the other hemisphere.

Individual sunspots may last a few hours to a few weeks. A spot group may persist for several months. It was by following sunspots across the solar surface that the rotational speed of the sun was first determined. Figure 2.17 shows how this velocity varies with latitude. As noted above, the sun does not rotate as a solid body. The rotational period is about 27 days at the equator, as observed from the earth, but rotates more slowly towards the poles.

Most of the interesting solar activity which affects earth-space revolves around magnetic activity which is indicated by sunspots, and the variation in their character. One of the most straightforward manifestations of this is the sunspot number, which reveals the 11 year solar cycle.

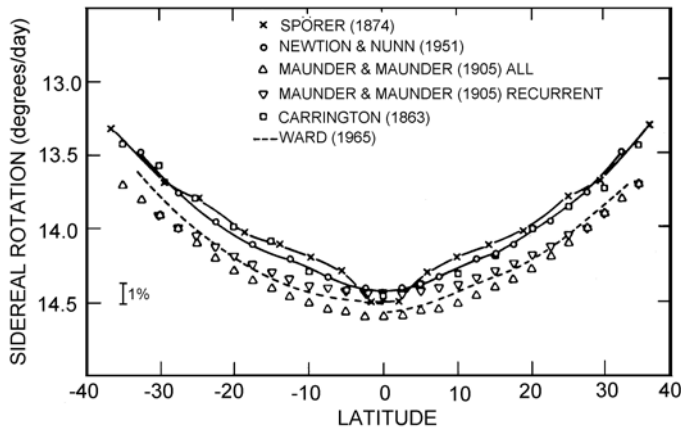


Figure 2.17 Period of the sun's rotation, as determined from sunspots. The results of Newton and Nunn [1951] refer to individual spots. The others refer to spot groups. From: Robert Howard, *The Rotation of the Sun, Reviews of Geophysics and Space Physics*, 16, 721-732, 1978. See also, Robert Howard, the Rotation of the Sun, *Scientific American*, page 106, 19?? (~1978). ($\tau \sim 25$ days at equator - appears longer from earth)

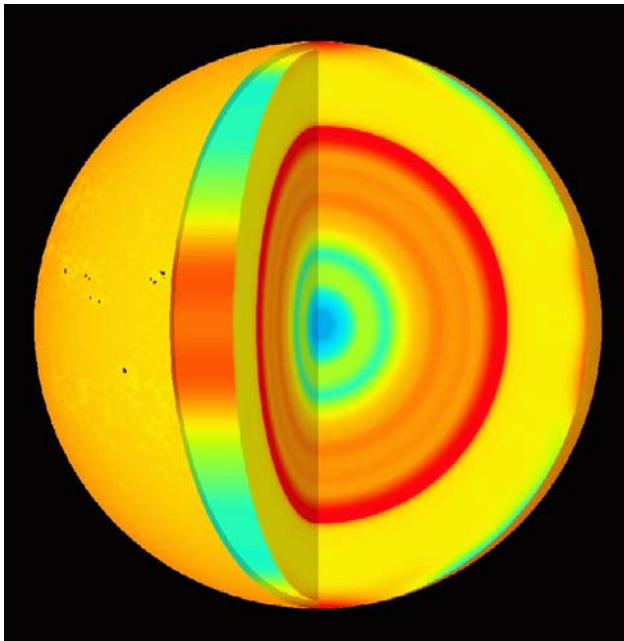


Figure 2.18 Solar rotation and polar flows of the Sun as deduced from measurements by MDI. The cutaway reveals rotation speed inside the Sun. The left side of the image represents the difference in rotation speed between various areas on the Sun. Red-yellow is faster than average and blue is slower than average. The light orange bands are zones that are moving slightly faster than their surroundings. The new SOHO observations indicate that these extend down approximately 20,000 km into the Sun. Sunspots, caused by disturbances in the solar magnetic field, tend to form at the edge of these bands.

3 Sunspot Number and the Solar Cycle

Statistically the sunspot number is closely correlated with many aspects of solar and geophysical activity. Sunspot number as defined below has proven to be a useful index of solar activity. In 1848 Rudolf Wolf in Zürich, Switzerland established the following index to characterize the "spottiness" on the solar surface:

$$R = h(10g + s)$$

where R = Wolf or Zürich sunspot number
 g = Number of sunspot groups (2 or more)
 s = Number of individual sun spots
 h = Subjective correction factor (Fudge factor)
 depending on instrumentation and observer.

A plot of monthly averaged sunspot numbers is shown in Fig 2.18 in which several features are apparent.

There is a period of about 11 years between consecutive maxima, but the cycle is not completely regular and varies in period from 8 to 15 years.

The rise time is about 4.8 years and decline is about 6.2 years.

The amplitude of the maximum can vary by almost a factor of 3. The largest annual mean number (190.2) occurred in 1957.

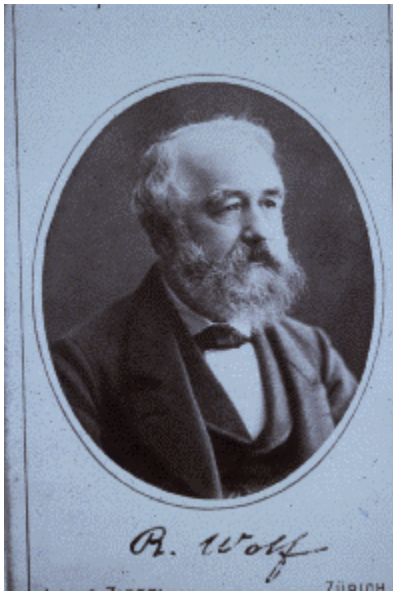


Figure 2.19a Johann Rudolph Wolf (1816-1893).

National Portrait Gallery,
Smithsonian Institution,
Washington D.C.

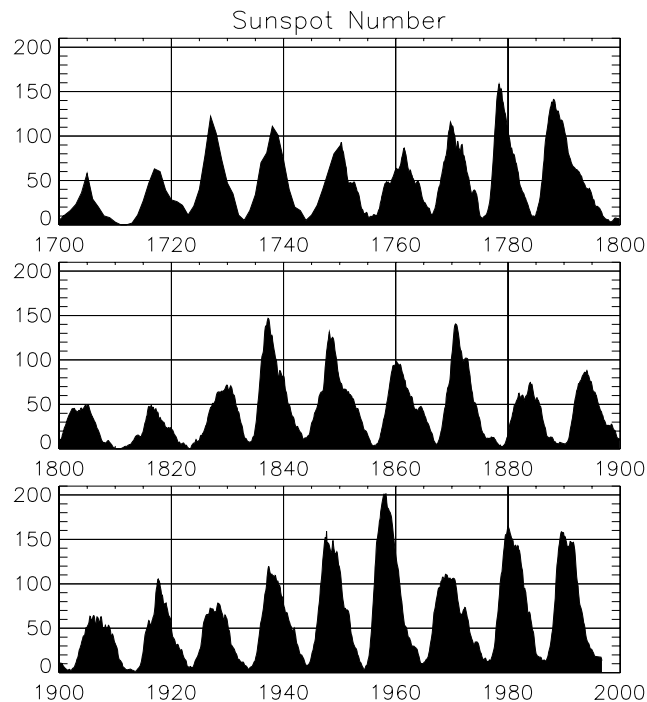


Figure 2.19b Monthly values for sunspot number. (Annual values used 1700-1750).

In 1848 Rudolph Wolf devised a daily method of estimating solar activity by counting the number of individual spots and groups of spots on the face of the sun. Wolf chose to compute his sunspot number by adding 10 times the number of groups to the total count of individual spots, because neither quantity alone completely captured the level of activity. Today, Wolf sunspot counts continue, since no other index of the sun's activity reaches into the past as far and as continuously. An avid astronomical historian and an unrivaled expert on sunspot lore, Wolf confirmed the existence of a cycle in sunspot numbers. He also more accurately determined the cycle's length to be 11.1 years by using early historical records. Wolf, who became director of the Zurich Observatory, discovered independently the coincidence of the sunspot cycle with disturbances in the earth's magnetic field.

<http://web.ngdc.noaa.gov/stp/SOLAR/SSN/ssn.html>

The latitudes where sunspots are formed changes in a systematic way during the 11 year cycle. At the start of a cycle (minimum) the spots tend to form near $\pm 35^\circ$ solar latitude. As the cycle progresses they form at successively lower latitudes as shown in Figure 2.20 down to about $\pm 5^\circ$ just as the next cycle begins at the higher latitudes. There is about a 2 year overlap between successive cycles when both high and low latitude spots appear.

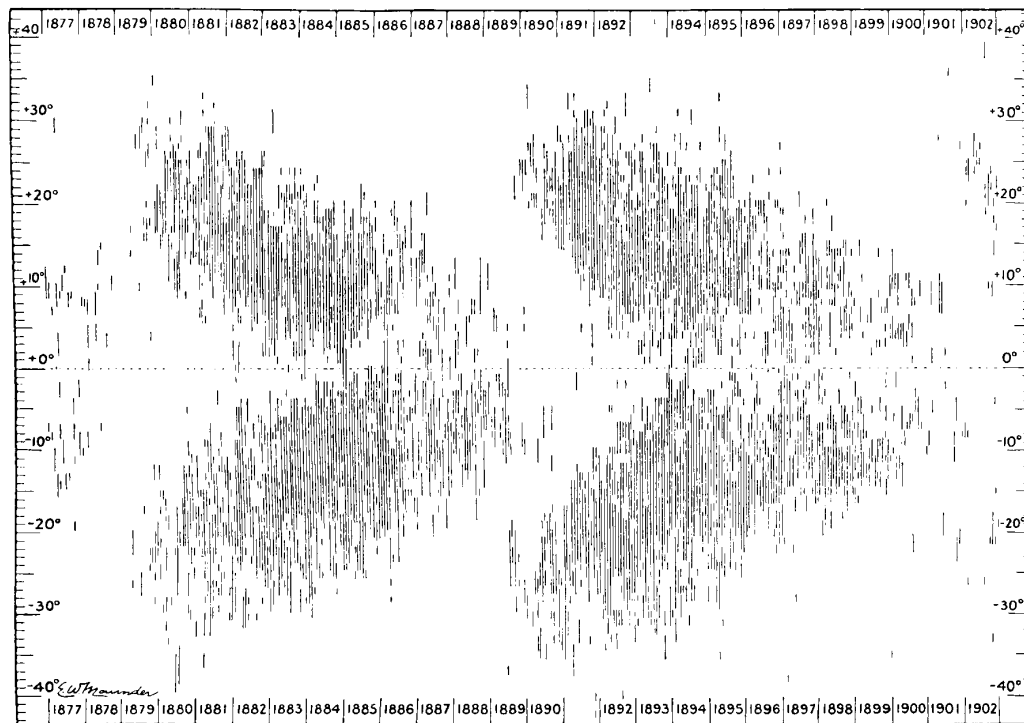


Figure 2.20 The Maunder "butterfly" diagram showing variation in distribution of sunspots with time. During a sunspot cycle the distribution migrates toward the solar equator. From: E. W. Maunder, "Notes on the Distribution of Sun-Spots in Heliographic Latitude, 1874-1902," *Royal Astronomical Society Monthly Notices*, volume 64, pages 747-761, 1904. (Found in the book by Edward Tufte, *Envisioning Information*, Graphics Press, Cheshire, CT., 1990). The vertical strips reflect the latitudinal range of a given measurement. For a modern version, see <http://www.ssl.msfc.nasa.gov/ssl/pad/solar/sunspots.htm>, and the butterfly diagram by David Hathaway at NASA/MSFC -<http://www.ssl.msfc.nasa.gov/ssl/pad/solar/images/bfly.gif>

For any given cycle the leader spot in any given group normally has the polarity of the nearer pole at the beginning of the cycle (minimum). The orientation of polarity remains unchanged during any given cycle. However, shortly after solar maximum the polarity of the large scale magnetic field weakens and reverses. When the next cycle starts (at solar minimum) spots formed at high latitudes will have the opposite polarity orientation compared to the old, low latitude spots belonging to the previous cycle. Thus the time required to return to the original magnetic configuration is about 22 years. For this reason, a complete solar cycle is typically considered to be 22 years, not 11 years.

The cause for much of this activity appears to have its source in the differential rotation of the sun's surface. As noted above, the sun does not rotate as a solid body. On earth the angular rotation rate is 360° per day regardless of latitude. On the sun the angular rotation rate at the equator is about 30% faster than at the poles. The rotation rate at the equator is about $13^\circ/\text{day}$ on the equator and about $10^\circ/\text{day}$ near the poles. As viewed from the earth, the sun's rotational period is generally considered to be ~ 27 days, and most solar related effects observed at earth show variations with this period. The general magnetic field is "frozen into" the surface due to the high plasma density. In the photosphere and below the kinetic energy density is much higher than the field energy density and hence the plasma motion determines the shape of the field.

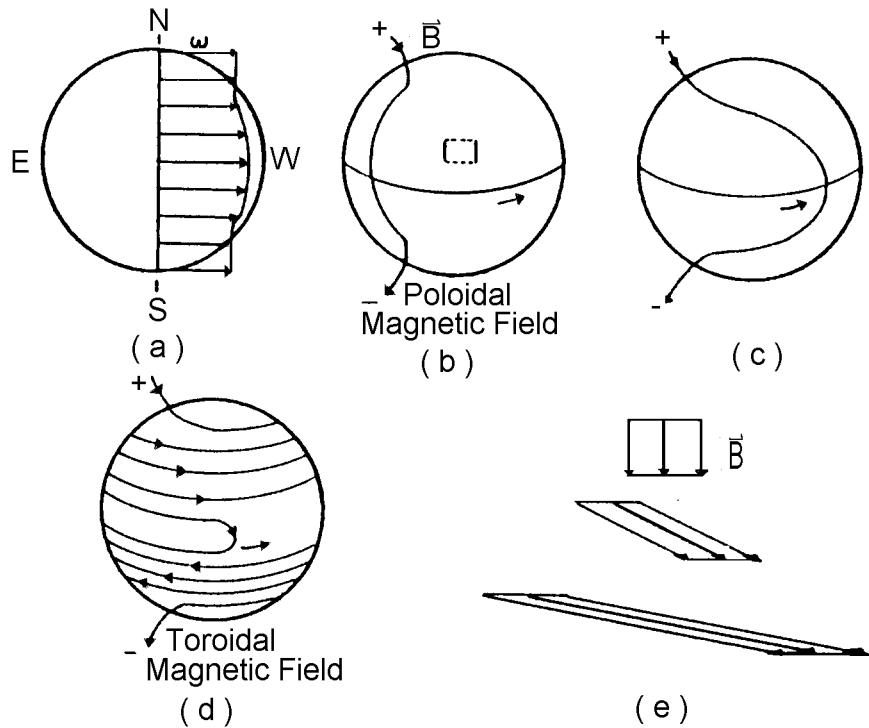


Figure 2.21. Enhancement of surface solar magnetic fields by differential rotation. (a) The general variation of angular velocity with latitude. Starting with a frozen-in longitudinal magnetic fields as shown in (b), differential rotation will wind it up as shown in (c) and (d). (e) The effect on the surface fluid element denoted in (b). The field lines are moved closer together under the shearing action of differential rotation and the field strength is thereby increased. (b to d are after Livingston, 1966. Gibson, 1972]

The differential rotation winds up the general magnetic field lines which initially lie in planes through the axis of rotation (a poloidal field). As this process proceeds, the field lines are eventually wrapped around the axis of rotation (a toroidal field) and the strength of the field is greatly increased as a result, as shown in Figure 2.21.

Convective motions below the photosphere further increase the density of the field lines twisting the tubes of enhanced flux into rope like structures (Figure 2.22 a). Kinks in these ropes of flux caused by small scale turbulence can produce even greater field strength within small local regions. When field strength (and hence the magnetic pressure) become large enough the flux tube can become buoyant and wells up through the surface. The regions where the tube penetrates the photosphere are known as sunspots and the field above fan out into a loop configuration with preceding (p) and following (f) spots of opposite polarity as shown in Fig (2.22 b).

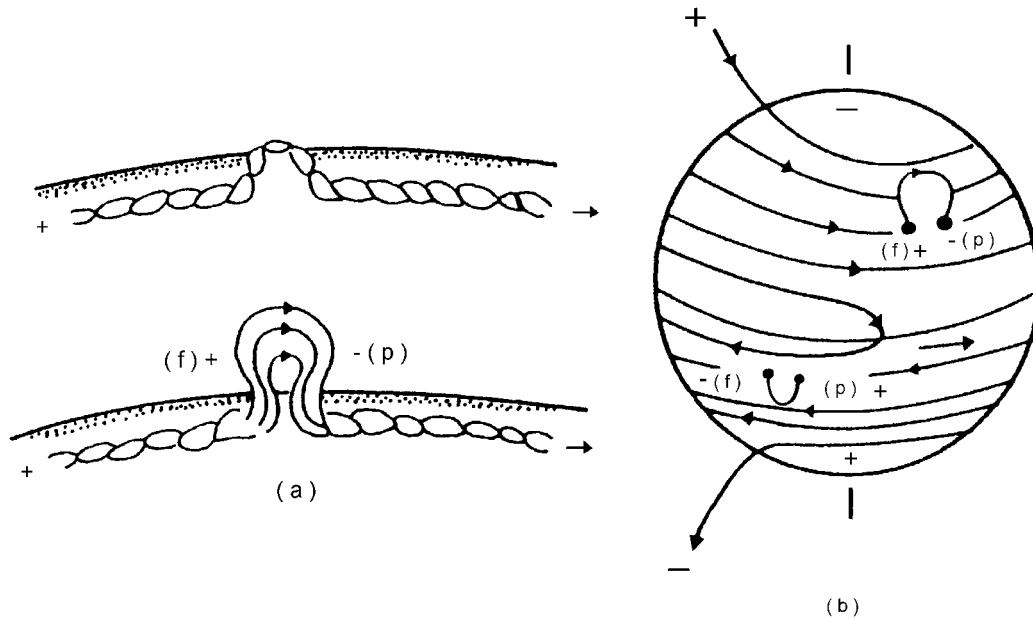


Figure 2.22 Formation of the bipolar sunspot magnetic field configuration. (a) Enhanced flux lines twisted by convective motion below the photosphere. (b) Spools formed by buoyant magnetic field (Gibson, 1972).

F Solar Activity

1 Prominences

Although the total energy output of the sun appears to be very stable there are a great variety of local disturbances discernible on the solar disk. In these so called "Active Regions" significant fluctuations in density, temperature and energy output can occur for periods lasting from minutes to weeks. The solar surface is a very turbulent place and many types of instabilities have been identified and should be studied in the photosphere, the chromosphere and the corona. The two examples of such behavior we will consider are prominences and solar flares.

Prominences, as illustrated in Figures 2.23 and 2.24, are extrusions of 'cold' chromospheric gas up into the corona, extending up out of active regions of enhanced magnetic field strength. They often show a loop like structure, as in the illustration, indicating that the plasma is confined by a loop like magnetic field structure, like that shown in Figure 2.23. When viewed against the sun's surface, they appear dark, since they are cooler than the background photosphere. Under these conditions, they acquire the name 'filament', but they are the same physical feature. Prominences are relatively long lived, often lasting many days. They are carefully watched by solar observers, as they often evolve into solar flares.

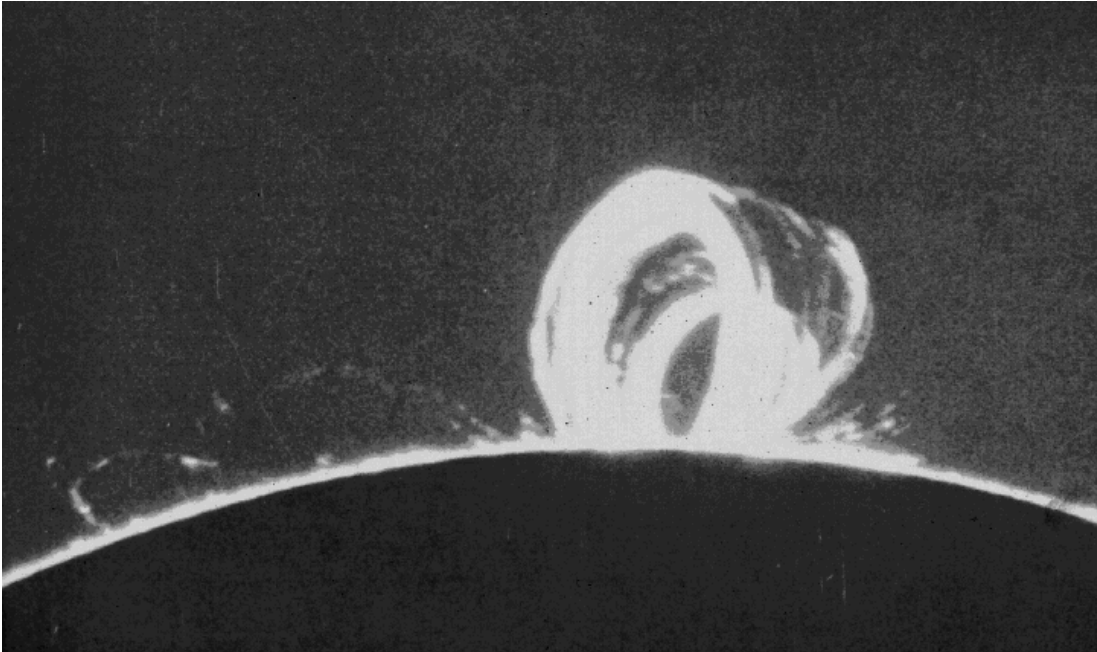


Figure 2.23 Loop Prominences, photographed in H-alpha. The image was taken circa June or July 28, 1957, taken at Sacramento Peak Observatory by Dr. Richard Dunn, a graduate student at that time. (<http://www.sunspot.noao.edu/gifs/loops.gif>) Since then, the (solar) Vacuum Tower Telescope he later created (1969) at the National Solar Observatory at Sacramento Peak, New Mexico was renamed as the Richard B. Dunn Solar Telescope. (1998)

Eruptive prominences, such as the one illustrated in Figure 2.24 can produce major disturbances in the corona, and solar wind (chapter 3). These structures emerge over a period of 1-2 hours. Still, the most significant changes in the near earth satellite environment come from solar flares. Solar flares emerge from magnetically active regions, typically where prominences and sunspot groups are also found. When these "kinks" in the field relax and the field lines begin to return to their original shape large amounts of magnetic energy become available in fairly small regions which turn "active" as a result.

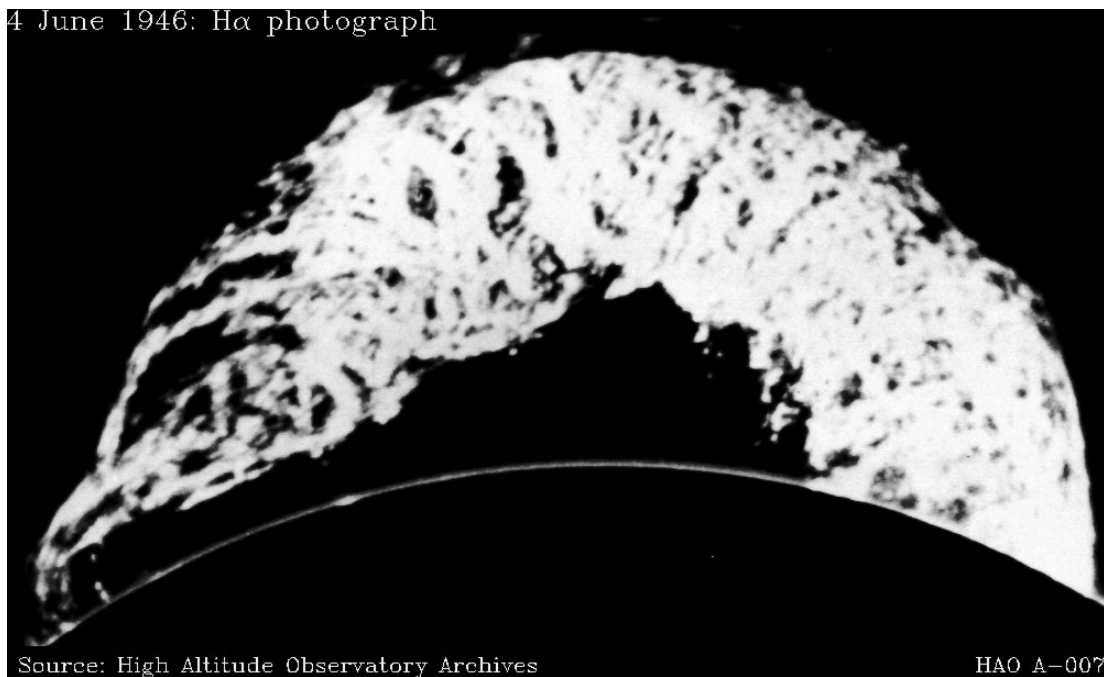


Figure 2.24 This solar eruption, an eruptive prominence, occurred on 4 June 1946, and is one of the largest ever recorded. It developed in 2 hours, reaching an altitude of nearly 1 million miles before disappearing.

From: The McGraw-Hill Encyclopedia of Space, page 506, 1968. Found on WWW, courtesy Paul Charbonneau, HAO. The progression of this famous prominence is found in: The Solar Atmosphere, Harold Zirin; The Sun, Our Star, Robert Noyes; Our Sun, Donald Menzel, and A Star Called the Sun, George Gamow.

2 Solar Flares

Prominences, as illustrated above, are dramatic but relatively common solar features. They consist of chromospheric plasma, that is plasma with a characteristic temperature of 4000-5000 degrees. (It is the fact that the contents of the prominence are cooler than the background solar surface which cause them to appear as dark filaments against the sun. They are only perceived as being bright in comparison to the tenuous corona). Solar flares, by contrast, are much less common, and very hot. These plasmas differ from those found in prominences, in that they are much more energetic. Viewed against the background of the sun, they are bright regions. They have even been observed by 'naked' eye observers, with a report by Carrington and Hodgson dated Sept 1, 1859 being one of the first such reports. They duly noted subsequent magnetic activity at earth, beginning the idea that there was a relationship between solar activity and earthly activities. (Phillips, Guide to the Sun, page 32) Figure 2.25 shows the famous "sea-horse" flare, which resulted in one of the largest magnetic storms observed during the space age. It resembles the prominence in shape, again reflecting the strong effect of the magnetic field structure on the plasma behavior.

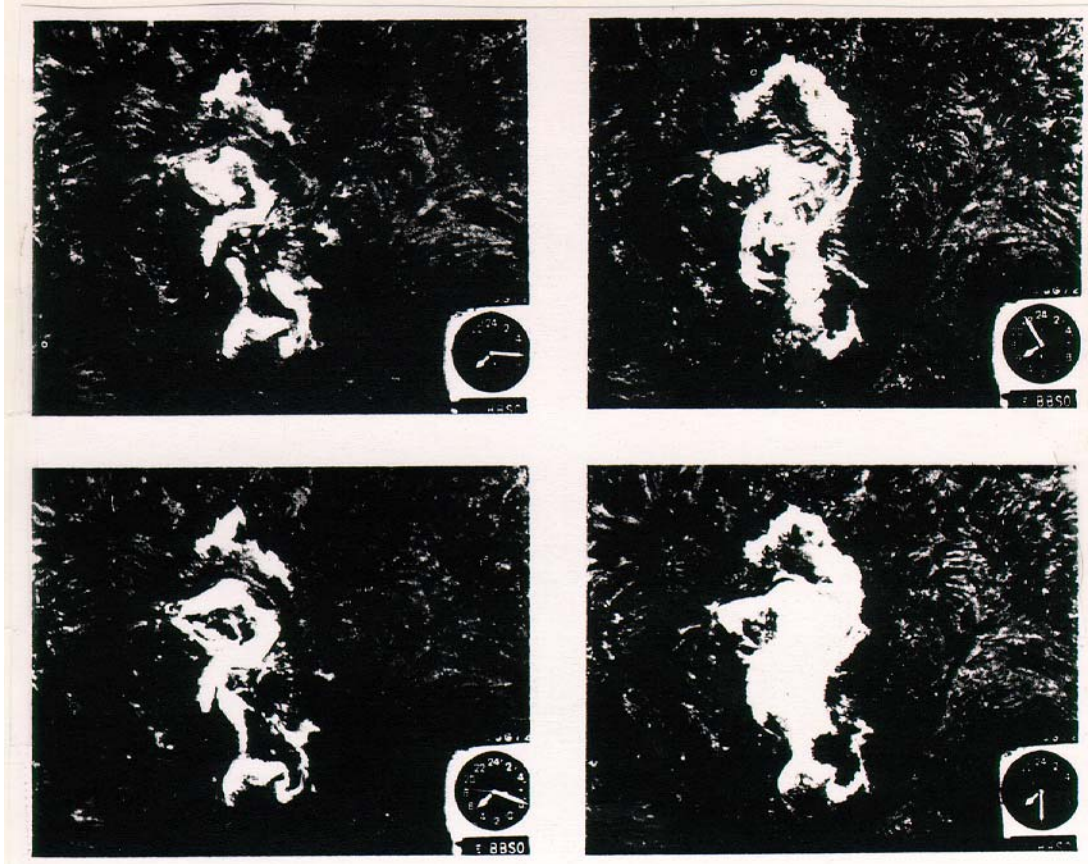


Figure 2.25 The development of a solar flare on 7 August 1972 is shown in sequence, in images taken at Big Bear Solar Observatory (BBSO). The progression starts at 1515, in the top left. The flare has visibly brightened by 1520 (bottom left), and has reached its peak brightness by 1530 (bottom right). By 1550 (top right), the flare has noticeably decayed. From: Cambridge Encyclopedia of Astronomy, edited by Simon Mitton, page 149, 1977.

Radiation of all frequencies from radio to x-rays as well as pulses of high energy particles (electrons, protons and some alpha particles) are emitted in a typical flare. Flares are monitored by the NOAA/GOES platforms, and watched for in x-rays and energetic particles (solar protons). Figure 2.26 shows data from a more sophisticated instrument on the Solar Maximum Mission, which observed X-rays. Figure 2.26 shows the 24-48 keV ($\sim 0.5\text{-}0.25 \text{ \AA}$) channel for 4 flares. The largest fortuitously came immediately after the SOLAR MAX rescue mission, on 24 April 1984. Notice the abrupt rise in flux, over a period of less than a minute. Energetic particles will start reaching the earth some hours later, and the plasma (shock wave) in the solar wind, several days later. The x-ray, and extreme ultraviolet (UV) flux has immediate effects on the upper atmosphere, heating it, and ultimately causing low altitude satellites to experience increased drag (Chapter 6). Note that the x-ray flux from one flare is greater than the x-ray production of the 'quiet' sun, integrated over the entire solar surface. High altitude satellites are more directly influenced by the solar wind shock wave arriving a few days later, while the energetic protons are primarily a problem for humans in space, but can lead to the degradation of electronics..

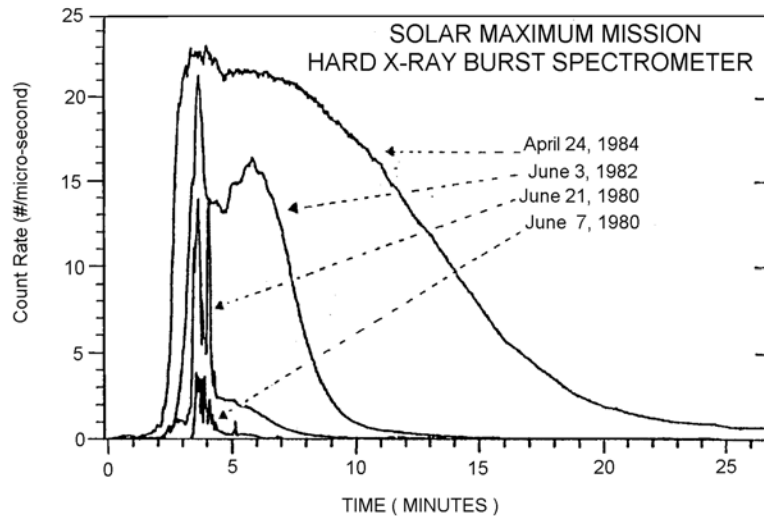


Figure 2.26 The development of 4 solar flares in hard x-rays. Figure from Dr. Alan Kiplinger, NASA/GSFC, March 1988.

While the visible emission from a flare increases by at most a few percent, x-ray emissions may increase as much as four orders of magnitude as measured by satellites above the earth's atmosphere (e.g. Figure 2.26). Measurements are made in several x-ray bands and flares are also classified according to the peak energy flux of the core. The increase in UV and higher energy photons is important for the degree of ionization of the ionosphere, and atmospheric heating as noted above. Also, it is relatively easy for geosynchronous weather satellites (e.g. GOES) to monitor solar activity via x-rays.

Flares are classified according to their size and intensity. The size or "importance" is based on flare area measured in millionths of solar disk area, and in the Doppler shift found in the H-alpha observations (e.g. the velocities). Duration is also directly related to the "importance". The total energy released during a flare may range from 10^{21} to 10^{25} Joules integrated over the three phases (1) precursor (or preflare) phase (2) flash and (3) main (or gradual) phase. The mean duration of a flare is roughly correlated to its magnitude (or importance)

IMPORTANCE	AVERAGE DURATION	PERCENT OF ALL FLARES
0	17 minutes	75
1	32 minutes	19
2	69 minutes	5
3 and 4	more than 2 hours	less than 1

Various models of solar flares have been proposed, but none explains all the flare observations. It is currently believed that the solar flare energy is probably stored in the twisted and kinked magnetic field lines above the active solar regions. By some triggering mechanism either within the field itself or from the outside the geometry of the field lines is rearranged and a portion of the stored energy is released. During this process large electric fields are generated which accelerate charged particles both outward and downward into the denser layers of the sun. It is estimated that between 10^{33} and 10^{36} electrons are accelerated to an average energy of 25 keV each second. These large fluxes of charged particles in turn are responsible for high levels of electromagnetic radiation emitted by the flare. Some fraction of the charged particles escape outward and eventually reach the earth where they are responsible for a number of effects which will be discussed in subsequent chapters.

There is an approximate empirical relationship between the number of flares per day and the sunspot number R defined above:

$$\text{number of flares/day} \sim R/25$$

which means one flare every few days at solar min when R is 5 - 10 and a maximum of several flares per day at solar max when R is 100 - 150, typically.

3 Coronal Mass Ejections

Coronal Mass Ejections (CME's) have conventionally been thought of as a manifestation of solar flares - that is a response in the upper solar atmosphere to the explosion occurring near the solar surface. There is now, however, a fairly vigorous debate in the solar-terrestrial community as to the causes of CME's, instigated to a certain extent by Jack Gosling, a researcher at Los Alamos National Laboratory. (See discussion in EOS, (Transactions of the American Geophysical Union), vol. 76, #41, page 401, 10 October 1995) The significance of the problem is that CME's are a major factor in magnetic storms on earth. Hence, understanding their origins is essential to understanding the causes of magnetic storms on earth.

They have been regularly monitored in recent times from ground based coronagraph instruments. Figure 2.27 shows further measurements from the satellite borne coronagraph on SMM, and a particularly well ordered mass ejection. Here, up to 10^{13} kg of coronal material may be ejected outwards at speeds as high as 1000 kilometers/second. In this case, a 'helmet streamer', as shown in the frame at 10:04, had been visible for a few days, during which time it showed little change in shape or brightness. An erupting prominence can be seen to form, and there may be material from the prominence visible in the bright filaments found at 13:10.

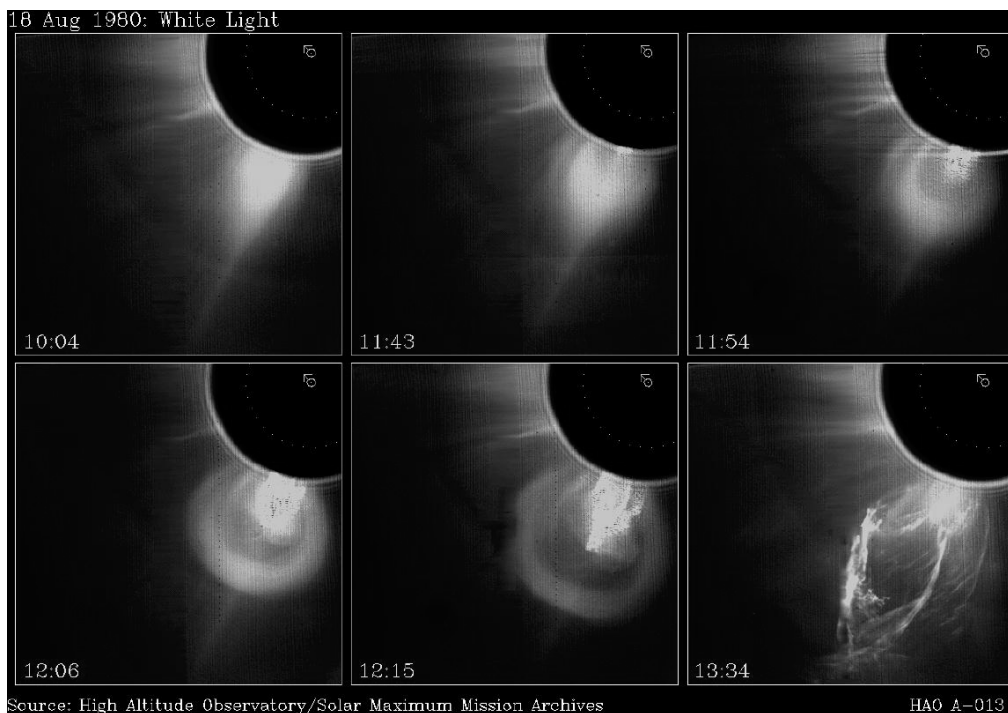


Figure 2.27 Coronal mass ejection, as observed by SMM, 18 August 1980. WWW, HAO, Charbonneau and White.

G Energy Production in the Sun

It is generally agreed that in the core of the sun ($r \leq 0.2 R_{\odot}$) energy is generated by the fusion of protons into He nuclei. The conditions for this self-sustaining reaction are brought about by the gravitational contraction of the sun which will compress and heat the core until thermonuclear ignition takes place.

Present calculations indicate that conditions in the core are

Density $\approx 10^5 \text{ kg/m}^3$ (about 100 times water density on earth)

Temperature $\approx 1.5 \times 10^7 \text{ K}$ (corresponds to an average energy of about 1 keV)

Pressure $\approx 2.5 \times 10^{11} \text{ atmospheres} \approx 2.5 \times 10^{16} \text{ (N/m}^2\text{)}$

There exist two major reaction cycles which convert atomic hydrogen nuclei into helium nuclei by thermonuclear fusion. In this process a small fraction of the mass of the system is converted into energy according to Einstein's equation

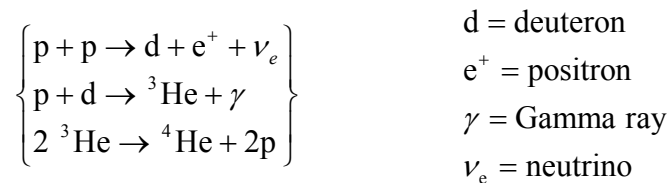
$$\Delta E = \Delta m c^2$$

where c = velocity of light = $3 \times 10^8 \text{ (m/s)}$

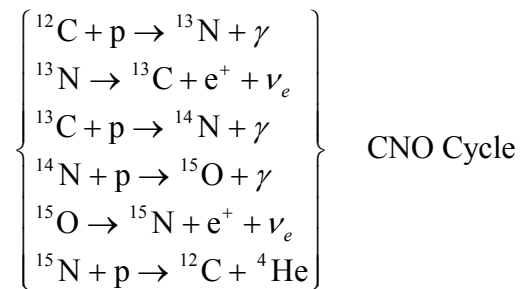
$\Delta m = 4(\text{proton masses}) - (\text{He}) \text{ mass}$

The loss of mass is about 0.7% of the original proton masses which means that about $4 \times 10^{-12} \text{ J}$ (25 MeV) of energy are generated per helium nucleus formed.

If the core temperature is actually about 15 million K most of the energy is generated by the so called proton-proton cycle



At slightly higher temperatures another cycle begins to function. This so-called CNO cycle goes something like this:



Note that in both of these cycles the net transformation is $4p \Rightarrow {}^4\text{He}$ plus energy and light particles such as e^+ , ν_e . Since neutrinos interact only very slightly with matter they will generally leave the sun carrying a certain amount of energy with them. For this reason the actual energy made available to the star is somewhat different for the two cycles.

Proton - Proton Cycle: 26.71 MeV/He nucleus formed

CNO Cycle: 25.03 MeV/He nucleus formed

Where $1 \text{ MeV} = 10^6 \text{ eV} = 1.6 \times 10^{-13} \text{ Joules}$. At the present rate of energy production the sun burns $6 \times 10^{11} \text{ kg}$ of hydrogen per second.

The extreme pressures and temperatures in the core of the sun are necessary to sustain these reactions for 2 reasons: (1) The particles must hit "hard enough" for the reactions to occur (high temperature) and (2) They have to hit often enough so that the energy emitted is high enough for the reaction to continue. Thus both high temperatures and high densities are essential. Finally we must have sufficient inward gravitational attraction to hold the whole thing together. Amazing!

H References

Bray, R. J., and R. E. Loughhead, Sunspots, Volume 7, International Astrophysics Series, Chapman and Hall, London, , 1964.

Durrant, C. J., The Atmosphere of the Sun, Adam Hilger imprint by IOP Publishing Limited, Bristol, England, 1988.

Eddy, John, A New Sun, The Solar Results from Skylab, NASA SP-402, Washington DC, 1979.

Friedman, Herbert, Sun and Earth, Scientific American Library, New York, 1986.

Gamow, George, A Star Called the Sun, The Viking Press, New York, 1964.

Gibson, E. G., The Quiet Sun, NASA SP-303., 1973.

Giovanelli, Ronald, Secrets of the Sun,

Hargreaves, J. K. (1979). The Upper Atmosphere and Solar Terrestrial Relations, Van Nostrand Reinhold, N.Y., Chapter 10.

Hufbauer, Karl, Exploring the Sun, Solar Science since Galileo, The Johns Hopkins University Press, Baltimore, 1991.

Lang, Kenneth R., Sun Earth and Sky, Springer Verlag, New York, 1995.

McIntosh, P. S. (1981). The Birth and Evolution of Sunspots: Observations, in The Physics of Sunspots (editors - L. E. Cram and J. H. Thomas), Sacramento Peak National Observatory, Sunspot, New Mexico, p. 7.

Menzel, Donald, Our Sun, Harvard University Press, Cambridge, Massachusetts, 1959.

Noyes, Robert, The Sun, Our Star, Harvard University Press, Cambridge, Mass., 1982.

Phillips, Kenneth, Guide to the Sun, Cambridge Press, Cambridge, 1992.

Zirin, Harold, The Solar Atmosphere, Blaisdell Publishing Co., Waltham, Mass., 1966.
Cambridge Encyclopedia of Astronomy, edited by Simon Mitton, Crown Publishers, NY, 1977.

I Problems

1. The temperature in the photosphere varies from 4400-6400 K. Calculate how the spectral peak varies in wavelength for the different regions (Wien's displacement law), and how the emitted power varies (Stefan-Boltzmann law).
2. During a solar eclipse we can see part of the solar atmosphere. Which part do we see and why?
3. Show that the photospheric pressure is only one percent of the earth's surface atmospheric pressure. Take the sun's surface gravitational field to be 30 times as great as on earth and the total mass of the solar atmosphere to be about 2.1×10^{19} kg compared to 5.29×10^{18} kg for the earth's atmospheric mass.
4. Assume that the average sunspot field is 0.3 Tesla within a volume of $(10^4 \text{ km})^3$. Compute the **percentage** decrease in stored magnetic energy needed to produce a flare with energy of 10^{25} Joules.
5. Plot n_e vs r for the corona, assuming a barometric relationship. The appropriate formula to use

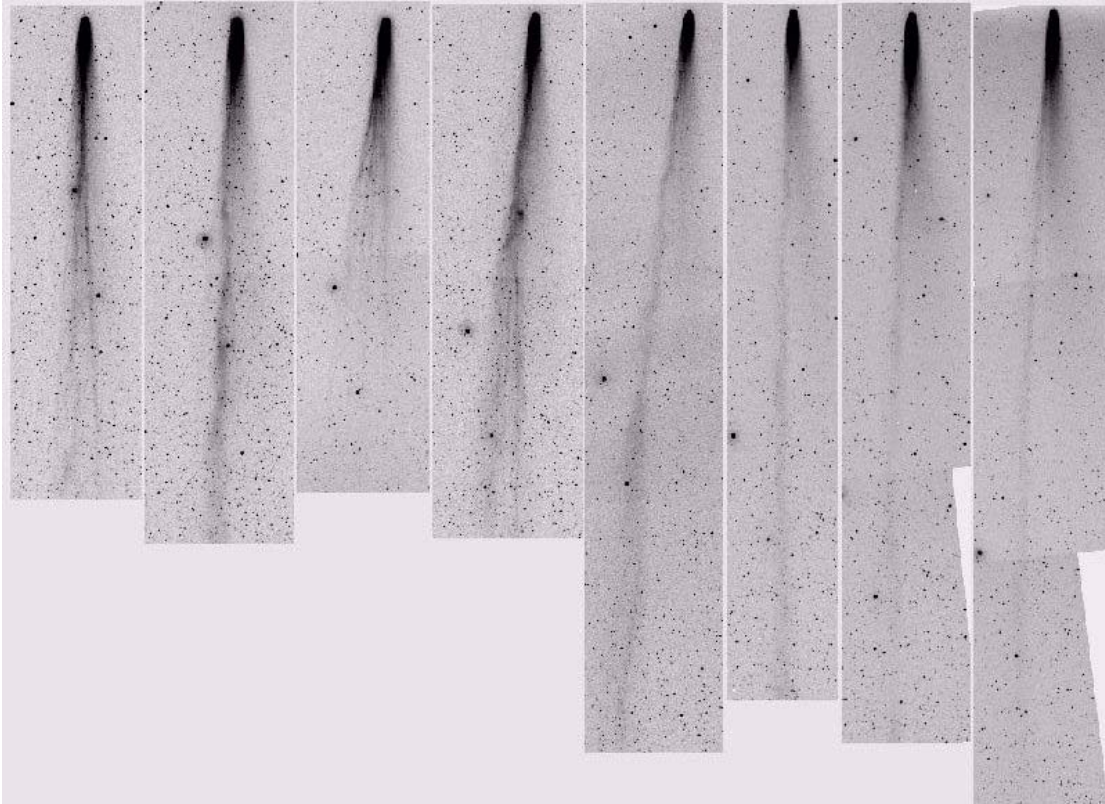
is:

$$n_e = n_0 \exp \left[\frac{G M_\odot \mu m_H}{kT R_\odot} \left(\frac{1}{r} - \frac{1}{r_0} \right) \right]$$

(From: *Astrophysical Formula*, Kenneth R Lang, Springer-Verlag, p 286, 1974.) Note that in this equation, the r and r_0 terms are in units of solar radii (e.g. 1, 2, 3...). μ is the effective mean molecular weight, which is ≈ 0.8 in the sun's atmosphere, taking into account the protons, alpha particles, and electrons. All the other numbers should be done in mks. Try temperatures of 1.0, 1.5, and 2.0 million K. Begin with $n_e \approx 3 \times 10^{14} \text{ m}^{-3}$ at $r = r_0 = 1$, and assume $n_e \approx 3 \times 10^{10} \text{ m}^{-3}$ at $r = 5$. What temperature produces the best result? Note that the barometric relation breaks down fairly quickly, but it produces one indirect indication of the temperatures which must exist in the stellar atmosphere.

6. The sun generates energy at the rate 4×10^{33} erg/sec by the conversion of hydrogen to helium. Show that if hydrogen could be completely converted to energy, it would require over 4 million metric tons of hydrogen to produce the observed solar energy generation rate. Assume that the energy conversion reaction is simply: $4\text{H}_1^1 \rightarrow \text{He}_2^4 + 2e^+ + \text{energy}$, where e^+ are positrons and the energy liberated per reaction is 25 MeV. (The energy comes off primarily in the form of neutrinos) Show that the sun would need to convert 6.68×10^{11} kg of hydrogen into 6.64×10^{11} kg of helium per second to produce the observed energy generation rate. (The mass defect for the energy conversion reaction is 0.0266 AMU).

This page intentionally left blank



Chapter 3 The Solar Wind

A Basic Characteristics

The solar wind is the supersonic flow of plasma produced in the upper reaches of the solar atmosphere. This flow of plasma outward from the sun defines most of our 'space weather' - and hence it is essential to understand before we can study the near-earth satellite environment. The solar wind was hypothesized to exist on the basis of a number of pre-space age observations, including observations of comets. Spectral observations of comets show that the dusty tail of a comet is pointed in a direction consistent with solar photon pressure, arcing away from the sun as one might predict. The plasma tail points directly away from the sun, however, and can show startling structure. Figure 3.1 shows a nice illustration of the plasma tail of a comet.



Figure 3.1 image of comet Hale-Bopp, taken on 1997 Apr. 4 (19h44-19h56UT) with 20-cm, f/2 Baker-Schmidt camera and Fujicolor 400 SG+ film. The field of view is about 5x3.5 deg. Copyright © 1997 by B. Kambic & H. Mikuz.
<http://www.fiz.uni-lj.si/astro/comets/images/95o1tcga-3.html>

The composition of the solar wind is similar to the composition of the sun's upper atmosphere ($\sim 90\%$ protons, $\sim 10\%$ He^{++}). The expansion speed is very slow in the inner corona, but increases rapidly outward, becoming supersonic at 10-20 solar radii. Here, supersonic means that the flow (or bulk) velocity of the solar wind is large compared to the average thermal velocity of the ions. (The ion mach number is greater than one). The characteristics of the solar wind at the earth's orbit (1AU from the sun) are given in Table 3.1.

Figure 3.2 illustrates most of the plasma properties at earth's orbit, for quiet times. The short term variability of the data seen here are fairly typical. Note that the electron bulk speed and ion speed should be very close - the electron measurement is a difficult one (by contrast to the ion measurement). Note the error bars on the electron measurement. Similarly, the proton and electron densities are relatively close, though one should keep in mind that quasi-neutrality only requires that the total ion charge density equal the electron density. Since there is a fair amount of He^{++} present, the proton density will nominally be about 90% of the electron density. By contrast, there is no overwhelming reason for the temperatures to be the same, though the variability in the values for ions again reflects the relative difficulty of the measurement - by contrast electron temperatures are fairly easy to obtain accurately.

Table 3.1

	Min Values	Average Values	Max Values
Flux (10^8 ions/cm ² sec)	1.0	3.0	10
Velocity (km/sec)	300	468	700
Density (ions/cm ³)	3.2	6.5	20
Electron Temperature (K)	9,000	140,000	2×10^5
Proton Temperature (K)	10,000	50,000	3×10^5
Magnetic Field Strength (nT)	2.2	6.6	10

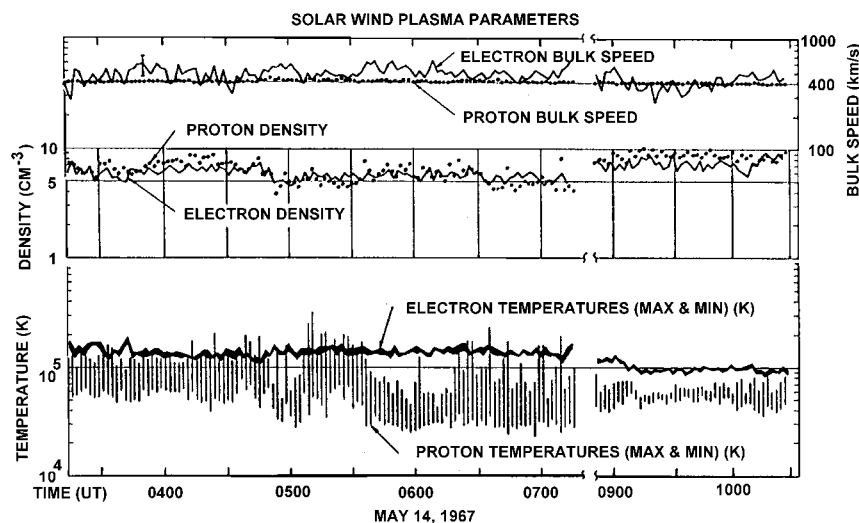


Figure 3.2 Solar wind plasma properties observed on Vela 4B

From: Montgomery, M. D., S. J. Bame, and A. J. Hundhausen, Solar Wind Electrons: Vela 4 measurements, *Journal of Geophysical Research*, 73, 4999, 1968.

The base values for the solar wind plasma illustrated above should not be considered all that typical. The variable nature of the solar wind plasma is illustrated by Figure 3.3, which shows a remarkable figure summarizing two decades of solar wind measurements, utilizing most of the data obtained during the space age. Data from 14 satellites are used, mapped to 1 AU (earth orbit). The data were smoothed using an 81 day running average, so fine scale events such as shock waves (section 3.5) are smoothed over.

These standard values will determine the standard size and shape of the earth's magnetosphere. Variations from these norms can result from solar activity, particularly flares. A major flare can cause a shock wave to be launched into the solar wind, doubling its velocity. In addition to the hot plasma, there are also more energetic particles in the solar wind. These energetic particles extend up in energy to an MeV or more.

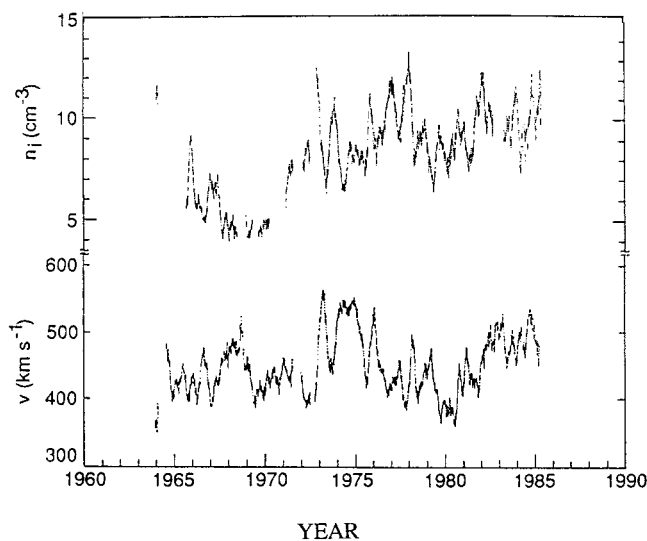


Figure 3.3 Solar wind ion density (top panel) and flow velocity (bottom panel) are shown for a 20 year period. Data are smoothed with an 81 day running average.

From: One year variations in the near earth solar wind ion density and bulk flow velocity, Scott J. Bolton, Geophysical Research Letters, vol. 17, p 37, 1990.

B The "Supersonic" Solar Wind

We begin with one of the first and most fundamental elements of the behavior of the solar wind. If the solar wind originates in the corona where the temperature is about 10^6 °K why is the solar wind velocity at 1 AU several times as high as the thermal velocities corresponding to the coronal temperatures? Figure 3.4 emphasizes this supersonic character of the solar wind. Charged particle data taken at 0.34 AU show ions with an energy/unit charge of about one keV, but the width (FWHM) is only a few tens of eV. The distributions are clearly supersonic.

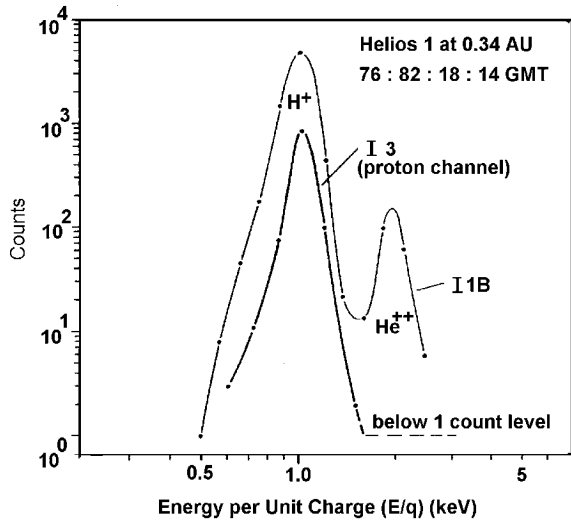


Figure 3.4 Count rate spectra simultaneously measured by the electrodynamic (mass analyzing) analyzer (I3) and electrostatic analyzer (I1B). The primary peak in both spectra is due to protons; the secondary peak in the electrostatic analyzer is due to He⁺⁺ (α) particles.

Rosenbauer, H., R. Schwenn, E. Marsch, B. Meyer, H. Miggenrieder, M. D. Montgomery, K. H. Mühlhäuser, W. Pilipp, W. Voges, and S. M. Zink, A Survey on Initial Results of the Helios Plasma Experiment, *Journal of Geophysics*, 42, 56-580, 1977.

The answer was obtained by Eugene Parker from the University of Chicago. He found that for a star like ours (e.g. one with the proper temperature) an acceleration of the outflowing gases to supersonic speed can occur. The process has been compared to the gas flow through a rocket nozzle (De Laval Nozzle) which also results in supersonic flow under the proper conditions. If the temperature of the corona is very low we have essentially a static atmosphere (such as the earth's) with very little escape of gas. The density in such an atmosphere is given by the barometric relation:

$$n(r) = n_0 \exp \left[\frac{G M_{\odot} m_H}{kT R_{\odot}} \left(\frac{1}{r} - \frac{1}{r_0} \right) \right] \quad (\text{Eqn. 3.1})$$

n is the plasma density at r ; n_0 is the plasma density at r_0 ;

r_0 is the reference altitude in solar radii, nominally 1 solar radius;

m_H is the mass of the proton (hydrogen nucleus) (in kg),

T is the temperature of the corona (in K or eV);

k is the Boltzmann constant; 1.38×10^{-23} J/K if T is in K, otherwise use $k = 1.6 \times 10^{-19}$ J/eV;

G is the gravitational constant,

M_{\odot} is the mass of the Sun (in kg); R_{\odot} is the radius of the sun (in meters)

If we now increase the temperature so that the atmospheric gases begin to escape into space we note that the total mass flow through any concentric sphere must be a constant (conservation of mass). We can write this as:

$$\rho(r) V_r 4 \pi r^2 = m n(r) V_r 4 \pi r^2 = \text{a constant} \quad (\text{Eqn. 3.2})$$

where V_r = radial velocity, $\rho = nm$ is the mass density.

We see from the form of $n(r)$, or $\rho(r)$, that it will decrease very sharply within the first few solar radii, depending on T , of course. (Homework problem 3.7). If the density falls off faster than $1/r^2$ (which it does) then V_r must increase to maintain constant mass flow. Thus we can achieve velocities higher than the speed of sound in the medium, or supersonic flow. If the temperature

gets very large then $\rho(r)$ is approximately constant and V_r has to decrease as we go further out. Thus very hot stars presumably show a stellar "breeze" whose speed decreases as we go outward. Once outside a distance of 5 or 10 solar radii, the acceleration process has ended, and the solar wind reaches an equilibrium velocity.

Figure 3.5 shows how the solar wind is observed to vary in density and flow velocity with distance from the sun. These data from 1978 were taken from Helios 1 and Voyager 1, from 0.3 to 1.0, and 1.0 to 5 AU, respectively. Over the observed range, the solar wind velocity is roughly constant at 400 km/s - the fluctuations seen here are a reflection of solar activity, as with figure 3.3.

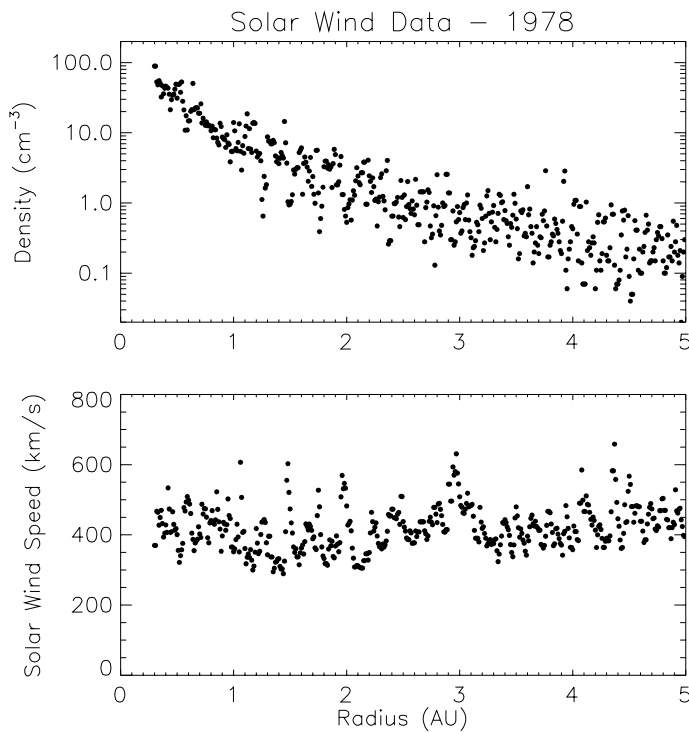


Figure 3.5 Solar Wind data from 1978, acquired from NSSDC for Helios 1 and Voyager 1.

C The Interplanetary Magnetic Field (IMF)

As we saw above, the plasma and the solar magnetic field form a closely coupled system. In the inner corona (out to about $3 R_{\odot}$) the magnetic field energy density dominates the kinetic energy density. This fact causes the inner corona to corotate with the sun like a rigid body (see Figure 3.6). This is possible since the magnetic fields can transfer angular momentum to the coronal plasma.

Beyond $3R_{\odot}$ the magnetic energy density falls off faster than the kinetic energy density of the plasma. At this point the outward moving plasma falls behind the sun's rotation and drags the magnetic field lines with it forming an Archimedes spiral structure. This process is somewhat analogous to the pattern formed by a rotating sprinkler and the angle between the radial direction and the field line is often referred to as the garden hose angle ψ (see Figure 3.6). At the orbit of the earth the garden hose angle ψ is about 45° and is given by

$$\tan \Psi = \omega_{\odot} r / V_{sw} \quad (\text{Eqn. 3.3})$$

where ω_{\odot} is the angular velocity of the sun (2.7×10^{-6} radians/second), r is the radial distance and V_{sw} is the speed of the solar wind particles. This formula was predicted by Eugene Parker (University of Chicago) before the advent of the space age, and the eventual confirmation of this relation is one of the great triumphs of space plasma theory. A simplified derivation follows here.

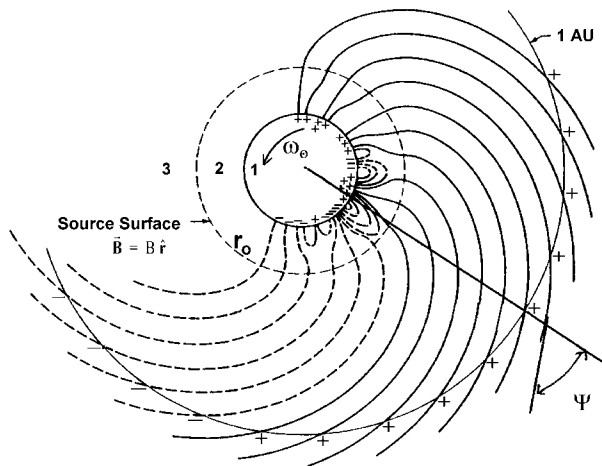


Figure 3.6 Spiral Structure of the solar wind. Three regions are defined here:

The solar surface, where the photospheric magnetic field can be obtained, as from observations at Mount Wilson.

The acceleration region, where the magnetic field can be calculated from potential theory

$$(\nabla^2 \phi = 0) \text{ (or } \nabla \times \mathbf{B} = 0)$$

Steady state region - magnetic field transported by solar wind as observed by

$$\text{satellites. } \frac{d\vec{B}}{dt} = -\vec{B}(\nabla \cdot \vec{V}) + (\vec{B} \cdot \nabla)\vec{V}$$

Figure modified from that found in Schatten, K.H., J. M. Wilcox, N. F. Ness, A model of interplanetary and coronal magnetic fields, Solar Physics, 6, pages 442-455, 1969.

We obtain our defining equation by considering a snapshot of the charges leaving the solar surface at some time t , and then integrate this equation in order to obtain the locus of charges in space. Note that this is not the trajectory of any individual charged particle. Just as in the water hose analogy, each particle is instantaneously moving radially away from the sun. Figure 3.7 illustrates the geometry to be used.

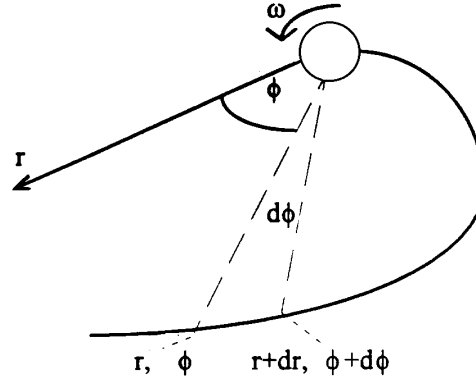
Figure 3.7 Geometry for IMF derivation

ω_{\odot} = angular velocity of the sun
 $r_0 = 3$ solar radii, the effective radius of the corona
 V_{sw} = solar wind velocity

Now:

Let r, ϕ be the coordinates of the particle at a time, t , and $r+dr, \phi+d\phi$ be the coordinates of particle 2 at the same time.

Let t_0 be the time at which particle 1 left the corona, t_0+dt_0 be the time at which particle 2 left the corona.



$$\text{Then: } t - t_0 = \frac{r - r_0}{V_{sw}}; \quad dt_0 = \frac{d\phi}{\omega_{\odot}}, \text{ and } t - (t_0 + dt_0) = \frac{r + dr - r_0}{V_{sw}}$$

$$\text{combining these we obtain: } dr = - \frac{V_{sw}}{\omega_{\odot}} d\phi$$

This can be integrated to obtain:

$$\int_{r_0}^r dr = - \frac{V_{sw}}{\omega_{\odot}} \int_{\phi_0}^{\phi} d\phi \quad (\text{Eqn. 3.4})$$

which is easily evaluated to be:

$$r - r_0 = - \frac{V_{sw}}{\omega_{\odot}} (\phi - \phi_0) \quad (\text{Eqn. 3.5}).$$

Here, r_0, ϕ_0 describes the position of the charge element at $t = 0$. We can make use of a fortuitous combination of numerical values, and then abandon mks units. Noting that the nominal value for the solar wind velocity is 400 km/s, we obtain:

$$\frac{V_{sw}}{\omega_{\odot}} = \frac{4 \times 10^5 \frac{m}{s}}{2.69 \times 10^{-6} \frac{radians}{s}} = 1.5 \times 10^{11} \text{ m} = 1 \text{ AU} \quad (\text{Eqn. 3.6})$$

Further noting that $r_0 \ll r$ for all observations made on satellites (r_0 is a few solar radii), and that we have the freedom to take ϕ_0 as zero, we obtain a very simple formula:

$$r (\text{AU}) = - \phi \quad (\text{Eqn. 3.7})$$

Referring back to Figure 3.7, we note that ϕ is measured in radians, and we are using the customary convention that the positive ϕ direction is counter-clockwise. Since r is positive, it is apparent that ϕ is a negative number - the spiral is pointed clockwise.

It can be shown that the radial component of the spiral IMF must decrease as r^{-2} , and we then have for the two components and magnitude of \mathbf{B} :

$$B_r = B_{r_0} \left(\frac{r_0}{r} \right)^2, \text{ and } B_\phi = -\frac{\omega_\odot r}{v_{sw}} B_r = -B_{r_0} \frac{\omega_\odot r}{v_{sw}} \left(\frac{r_0}{r} \right)^2 \quad (\text{Eqn. 3.8})$$

$$B = \sqrt{B_r^2 + B_\phi^2} = B_{r_0} \left(\frac{r_0}{r} \right)^2 \sqrt{1 + \frac{\omega_\odot^2 r^2}{v_{sw}^2}} \quad (\text{Eqn. 3.9})$$

The agreement of this model with in situ measurements is displayed in Figures 3.8 and 3.9, respectively.

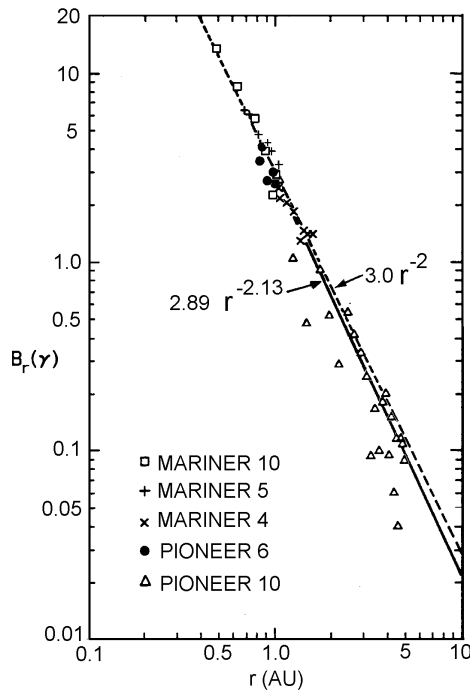


Figure 3.8 (left) Solar rotation averages of the magnitude of the IMF radial component B_r measured by Mariner 4, 5, and 10 and Pioneer 6 and 10. Curves showing and r^{-2} radial distance dependence (dashed curve) and the 'best' least squares fit to the combined data (solid curve) are included.

From: Behannon, K., Heliocentric distance dependence of the Interplanetary Magnetic Field, Reviews of Geophysics and Space Physics, **16**, 125-144, 1978

Magnetic field variation with distance from the sun

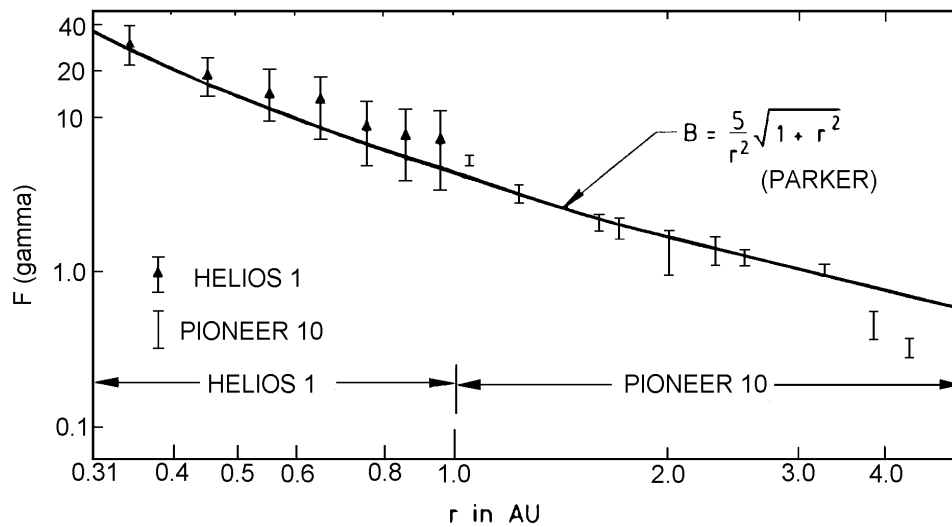


Figure 3.9 Range and central value (black triangle) of magnetic flux (F) in gamma ($1 \gamma = 1$ nano-Tesla, or 1 nT) for distance bins of 0.1 AU versus distance from the sun. Data are from Helios-1 and Pioneer 10. The fitted curve is from Parker's theory. G. Musmann, F. M. Neubauer, and E. Lammers, Radial Variation of the Interplanetary Magnetic Field between 0.3 and 1.0 AU , Observations by the Helios-1 Spacecraft, *Journal of Geophysics*, 42, 591-598, 1977

Finally, we see from figure 3.6 that the garden hose angle ψ is given by:

$$\tan \Psi = \frac{B_{\phi}}{B_r} = \frac{\omega_{\odot} r}{v_{sw}} \quad (\text{Eqn. 3.10})$$

as given at the beginning of this section. At 1 AU , $\tan \psi \approx 1$, and $\psi \approx 45^\circ$. This latter value is verified by observations, as shown in Figure 3.10. This summary of a decade of observations was obtained from the National Space Science Data Center (NASA/NSSDC) archives. Hourly observations were summarized here, binned in one degree increments. The figure indicates a peculiar problem, however. The expected observations are occurring at $\psi=45^\circ$, but there is a second peak at 225° . The former is due to magnetic fields pointed outward from the sun, the latter to inward pointing fields. Parker's theory is silent on this problem - the distinction is in the source region on the sun's surface. Why would there be such a variation?

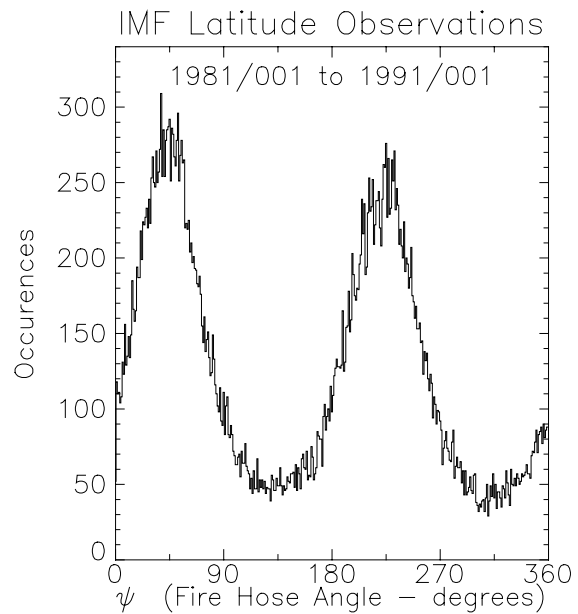
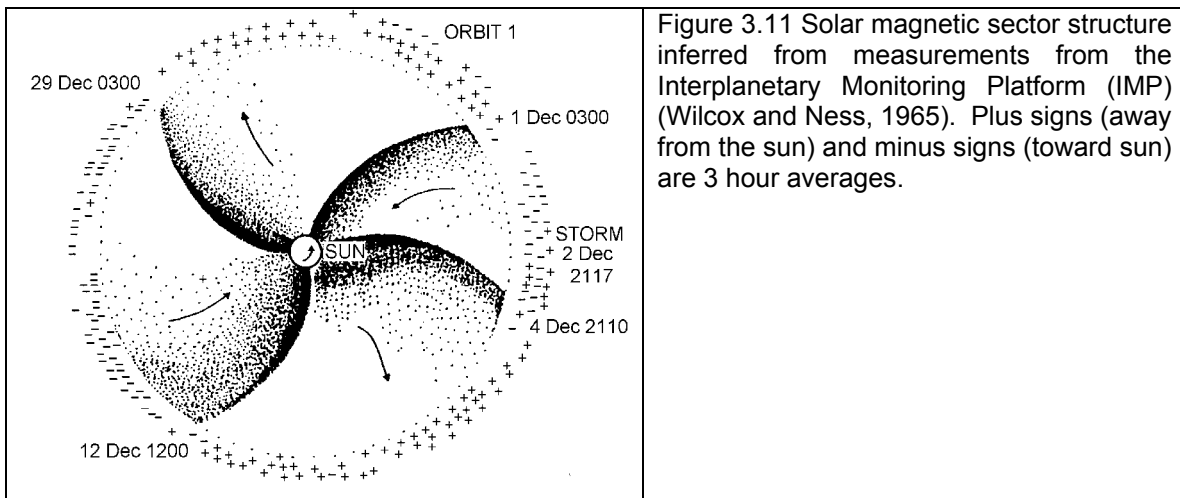


Figure 3.10 Solar Wind magnetic field data taken at earth orbit from 1981-1990.

D Sector Structure

The quandary noted above regarding the direction of the IMF was noted early in IMF field observations. Figure 3.11 shows a summary of a few early orbits of the IMP platform. The inward/outward duality was noticed to organize itself by solar rotation. Near the solar equatorial plane the interplanetary magnetic field (IMF) is organized into sectors or regions (typically four to six) where the magnetic field is predominantly directed either away from or toward the sun as shown in Figure 3.11. It was initially thought this structure reflected the longitudinal structure of the sun's surface field. It is now believed that this observational pattern reflects the relative motion of the earth, and its satellites, above and below the magnetic field lines which map to the sun's magnetic equator.



This sector structure is thought of as the result of the earth (effectively) moving above and below the sun's magnetic equator as the sun rotates. The 'heliospheric current sheet' is the boundary between the magnetic hemispheres of the sun. This boundary is uneven, and has been described (by Hannes Alfvén) as somewhat like the ruffled skirt of a ballerina. As this structure sweeps past the earth, the earth moves above and below the plasma originating from the sun's magnetic equator. This is the origin of the dual values for the 'garden hose' angle.

The current sheet is a natural consequence of having oppositely directed magnetic fields in the two solar hemispheres - a similar feature will be found in the earth's magnetosphere. This interpretation was largely proven correct by the recent Ulysses mission, which saw the disappearance of the sector structure as the satellite left the ecliptic plane, and circled the sun in a polar orbit; (E. J. Smith et al, Disappearance of the Heliospheric Sector Structure at Ulysses, Geophysical Research Letters, 20, page 2327, November 5, 1993)

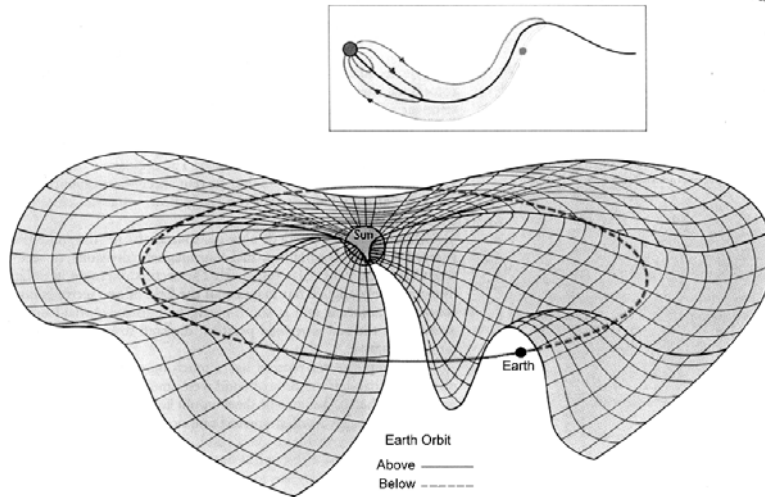


Figure 3.12 The Heliospheric current sheet.

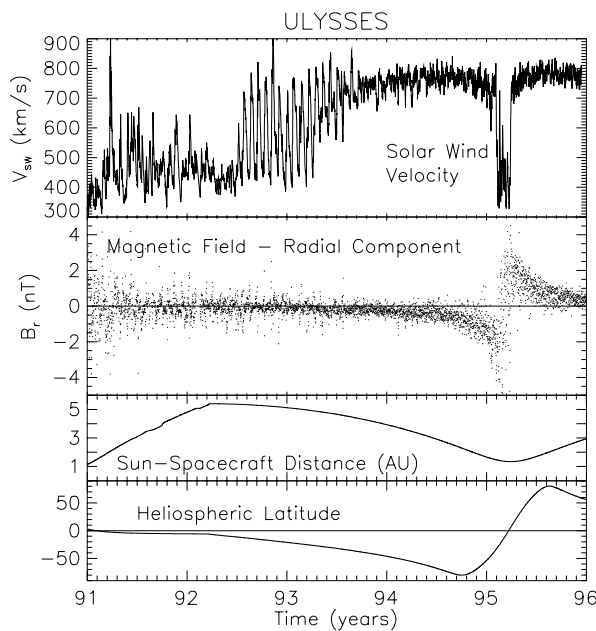


Figure 3.13. Summary of 5 years of data from the Ulysses SWOOPS instrument, obtained from the European Space Agency. The satellite cruised relatively slowly out to Jupiter (at 5 AU), then turned south and out of the ecliptic plane. Nearly four years after launch by the space shuttle Discovery, on 13 September 1994, the European-built spacecraft reached the most southerly point on its out-of-ecliptic orbit, 80.2 degrees south of the Sun's equator, at a distance of 2.3 astronomical units (345 million km) from the Sun. Less than a year later, the satellite passed over the north solar pole, reaching a peak latitude of 80.2°.

Note that above ~40° solar magnetic latitude, only high speed solar wind is observed, and that as the satellite moved below the ecliptic plane, primarily negative values of B_r were found. In 1995, at northern latitudes, positive values are found.

Data obtained from the European Space Agency (ESA) Ulysses archive. See URL <http://helio.estec.esa.nl/ulysses/>. The above result was first published by: J. L. Phillips, A. Balogh, S. J. Bame, B. E. Goldstein, J. T. Gosling, J. T. Hoeksema, D. J. McComas, M. Neugebauer, N. R. Sheely, and Y.-M. Wang, Ulysses at 50° south: Constant immersion in the high-speed solar wind, *Geophysical Research Letters*, 21, page 1105, 1994. Also, Ulysses Explores The South Pole Of The Sun, R.G. Marsden, ESA Bulletin No.82 May 1995)

E The Source of the Fast Solar Wind

The reality of the solar wind origin ended up being somewhat more complex than Parker's theory, particularly because of the complex magnetic field structure near the solar surface. Frequently, the solar wind speed would be observed to increase dramatically to ~ 800 km/s. The so-called "fast" solar wind streams were ultimately recognized as being a high-solar latitude phenomena, as discussed above. The evolution of observations and theory currently indicate that the solar wind must originate from regions called coronal holes – that is regions which appear dark in extreme ultraviolet (EUV) and x-ray images. In 1999 this general understanding was dramatically changed by the work of Hassler et al, illustrated below. Using the SOHO satellite imagers, they were able to deduce that the specific origin was at the boundaries defined by the chromospheric network, corresponding roughly to the edges of the cells defined by super-granulation.

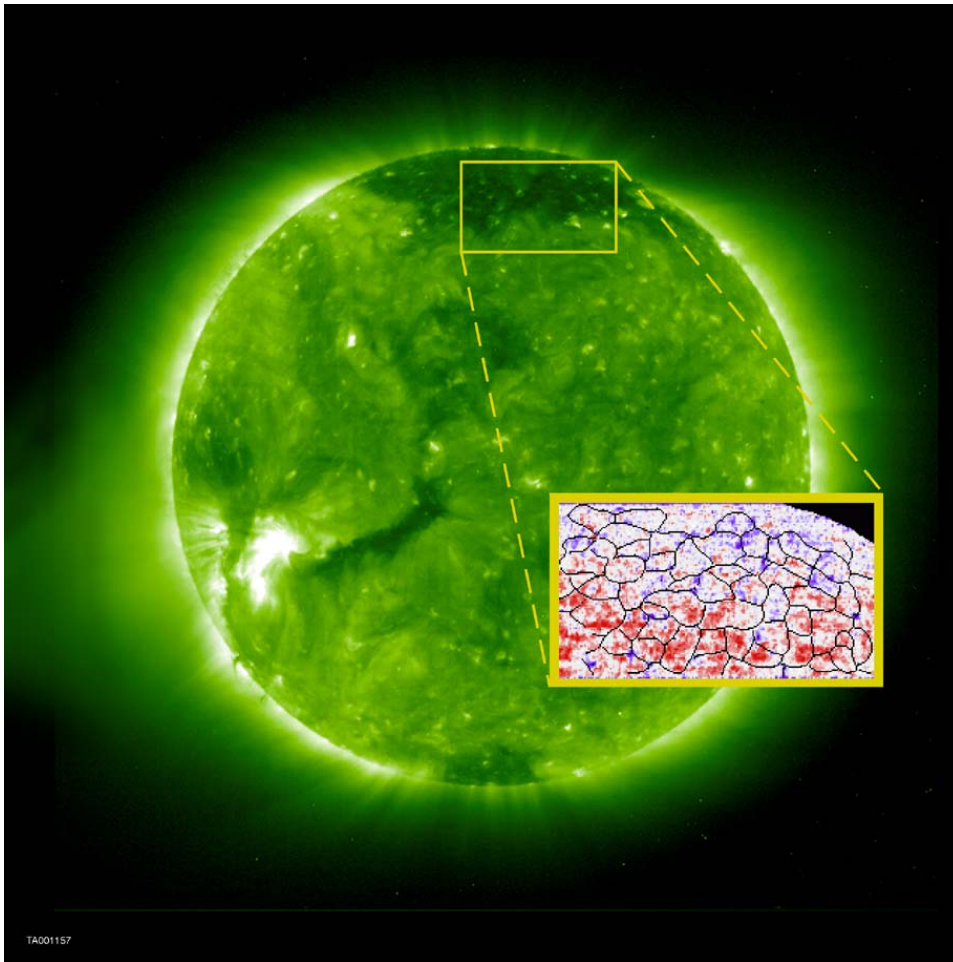


Figure 3.14 Extreme ultraviolet image of the Sun taken with the EAS/NASA Solar and Heliospheric Observatory (SOHO) Spacecraft revealing gas at 1.5 million degrees shaped by magnetic fields. Bright regions indicate hot, dense plasma loops with strong magnetic fields, while dark regions imply an open magnetic field geometry (coronal hole), and are the source of the high speed solar wind. Hassler et al, Solar Wind Outflow and the Chromospheric Magnetic Network, *Science*, 283, page 810, February 5, 1999. Photo Credit: ESA/NASA

http://www.boulder.swri.edu/~hassler/SW_press_release.html

Figure 3.14 shows a solar image from 22 September 1996 (near solar minimum) taken at 195 Å, imaging Fe XII at an approximate temperature of 1.5 million K. The "zoomed-in" or "close-up" region shows a Doppler velocity map of million degree gas at the base of the solar atmosphere, where the solar wind originates. This image is obtained from SUMER Ne VIII 770 Å line-of-sight Doppler velocity measurements, on a scale ranging from ± 10 km/s. Blue represents blue shifts or outflows and red represents red shifts or downflows. The atmospheric motion toward us, away from the solar surface, is seen as a blue shift, and is the beginning of the solar wind. The blue regions are inside the coronal hole, or open magnetic field region, where the high speed solar wind is accelerated. Superposed at the edges of the cells is the chromospheric network, where the strongest flows (dark blue) occur.

F Interplanetary Shocks

The largest fluctuations in the solar wind are the solar flare induced shock waves. Interplanetary shocks are responsible for magnetic storms on the earth, and ultimately, major anomalies in satellite operations. The explosion of plasma in the lower corona termed the solar flare, launches a shock wave which propagates at 800-1000 km/s outward into the solar wind. (Note, following the discussion at the end of chapter 2, it is perhaps most accurate to relate the shock wave to coronal mass ejections (CME's). There still seems to be a majority opinion that CME's are due to flares, however.)

Figure 3.14 shows a cartoon view of a shock. There is a shock front, which consists of hotter, denser plasma, with enhanced magnetic fields. Behind this shock front is the driver gas, which may have enhanced levels of alpha particles (He^{++}). This latter feature is thought to be a significant clue as to the nature of the flare process.

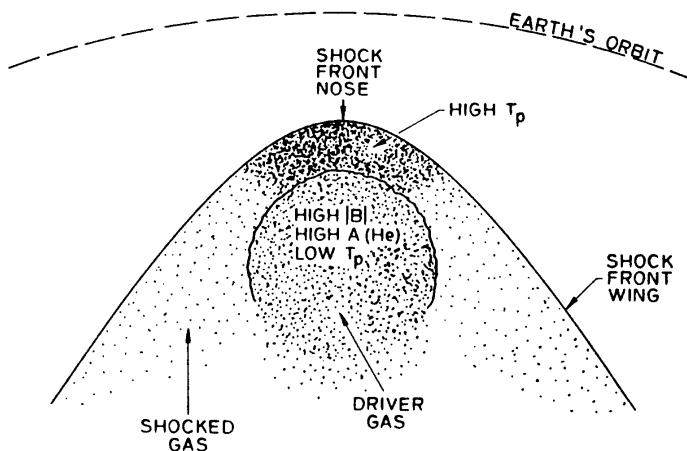


Figure 3.14 Cartoon for interplanetary shock approaching earth orbit. From: An analysis of shock wave disturbances observed at 1AU from 1971 through 1978, G. Borrini, J. T. Gosling, S. J. Bame, and W. C. Feldman, Journal of Geophysical Research, vol. 87, page 4365, 1982.

A sense of the kinds of changes that normally occur is given by Figure 3.15a, which shows an averaged view of 103 shocks observed by IMP6, 7, and 8, at 1 AU, from 1971 to 1978. The shock front takes less than a day to pass the earth; this is the period during which the magnetosphere will be most dramatically affected. Figure 3.15b shows a complementary view of the same process, using Pioneer 11 data, and a much finer time scale. It can be seen that the

leading edge transition is very sharp, and in this case the shock front passed the satellite fairly quickly (several AU downstream from earth).

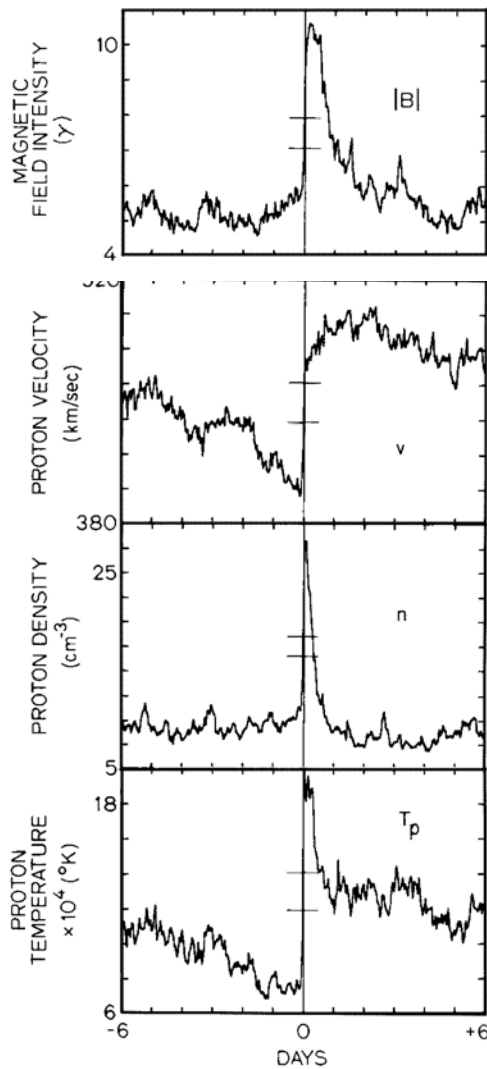


Figure 3.15a Superposed epoch plots of solar wind proton velocity, density, temperature, and magnetic field intensity, for 103 shock events at 1 AU. One-hour averages of data are employed. The small horizontal error bars indicate the normal variance of each parameter.

From: An analysis of shock wave disturbances observed at 1AU from 1971 through 1978, G. Borini, J. T. Gosling, S. J. Bame, and W. C. Feldman, *Journal of Geophysical Research*, vol. 87, page 4365, 1982.

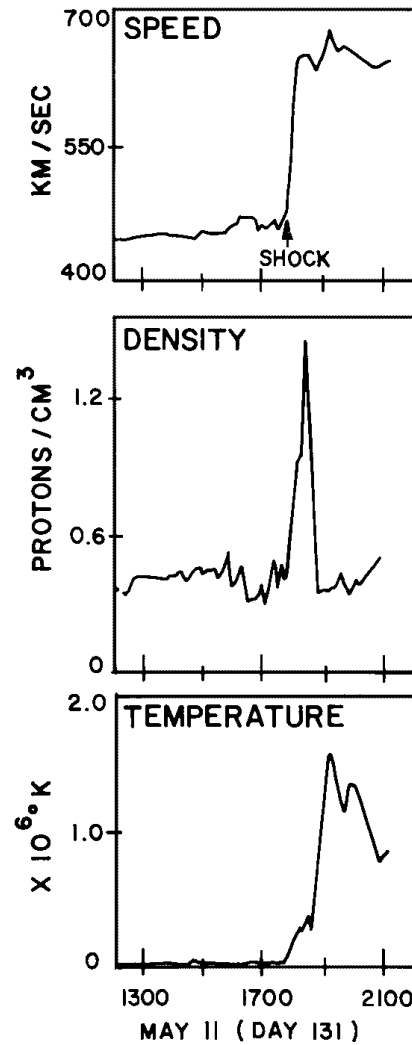


Figure 3.15b Plasma data from Pioneer 11, on May 11, 1978.

From: Plasma shocks and energetic particles in the outer solar system: Trapping and Asymmetry observations from Pioneer 10 and Pioneer 11, D. S. Intrilligator and W. D. Miller, *Journal of Geophysical Research*, vol. 87, page 4354, 1982.

G References

Akasofu, S. I. Interaction between a Magnetized Plasma Flow and a Strongly Magnetized Celestial Body with an Ionized Atmosphere: Energetics of the Magnetosphere, Annual Review of Astronomy and Astrophysics, Vol. 20, p 117, 1982.

Falthammer, C. G.. The Solar Wind, in Cosmical Geophysics, Egeland et. al. eds. Universitetsforlaget, Oslo Sweden, Chapter 7, 1973.

Hill, Thomas W. and Wolf, Richard A., Solar Wind Interactions in The Upper Atmosphere and Magnetosphere. National Sciences, Washington, D. C., p 25, 1977.

Hollweg, J. V. The Energy Balance of the Solar Wind, in The Sun as a Star, NASA Monograph, NASA SP-450., 1981.

Hundhausen, A. J., Coronal Expansion and Solar Wind. Springer-Verlag, N. Y., 1972.

Kopp, R. A. (1981) Heating and Acceleration of the Solar Wind, in The Sun as Star, NASA Monograph, NASA SP-450.

Wilcox, J. M. and Ness, N. F. (1965). Quasi-Stationary Corotating Structure in the Interplanetary Medium. J. Geophys. Res., 70, 5793.

H Problems

1. Calculate Beta (β) for the solar wind. Note that you need to modify the form used earlier just a bit, since this is a streaming plasma, with a velocity v :

$$\beta = \frac{\text{Kinetic Energy Density}}{\text{Magnetic Energy Density}} = \frac{\frac{1}{2}nmv^2}{\frac{B^2}{2\mu_0}}$$

2. Suppose, using the new space telescope we detect that a nearby hot star has a stellar wind flowing at 600 km/sec. Could we say that this stellar wind is supersonic? Explain.
3. If the solar wind speed doubles, how will the spiral (or "garden-hose angle") change? Does this situation ever occur? Explain.
4. Compare the solar wind particle density at the time of solar minimum with the particle density at solar maximum. Does the solar wind speed change during the solar cycle? Explain.
5. What does it mean to say that the "solar wind is supersonic"? Calculate ratio of the ion flow velocity to the average kinetic (thermal) energy of a 7 eV proton. This is the (ion) Mach number. Repeat for the solar wind electrons ($T \sim 10$ eV).
6. Waves can travel along magnetic field lines much like waves travel along a stretched string. The speed of these "magnetic" waves (called Alfven waves) is given by

$$V_A = \sqrt{\frac{B^2}{\rho\mu_0}}$$

where B is the magnetic field strength, ρ is the plasma density, and μ_0 is the magnetic permeability. Is the solar wind super-Alfvenic (i.e. faster than V_A)? If so, why is this important? Is this in any way analogous to being supersonic?

7. Use equations 3.1 and 3.2 to obtain V_{sw} as a function of r for the region from near the sun's surface out to $2 R_\odot$, for 3 possible temperatures. Take m = hydrogen mass (1 amu), T = 1,000,000 K, 1,500,000 and 2,000,000 K (1.0, 1.5, and 3.0 million degrees), $\rho = 10^{-10}$ gm / cm³ at $r = R_\odot$, and $V_o = 1$ km/s. The solar wind reaches supersonic velocity in this region. Show the mach number is greater than 1. ($M > 1$) at $2 R_\odot$, for the solutions you generate.

$$(M = \sqrt{\frac{\frac{1}{2}m v_{sw}^2}{\frac{3}{2}kT}} = v_{sw} / \sqrt{3kT/m} = v_{sw} / v_{thermal})$$

Figure 3.16 shows the complete solution to the expansion problem as formulated by Parker. (Your result will not look like this.)

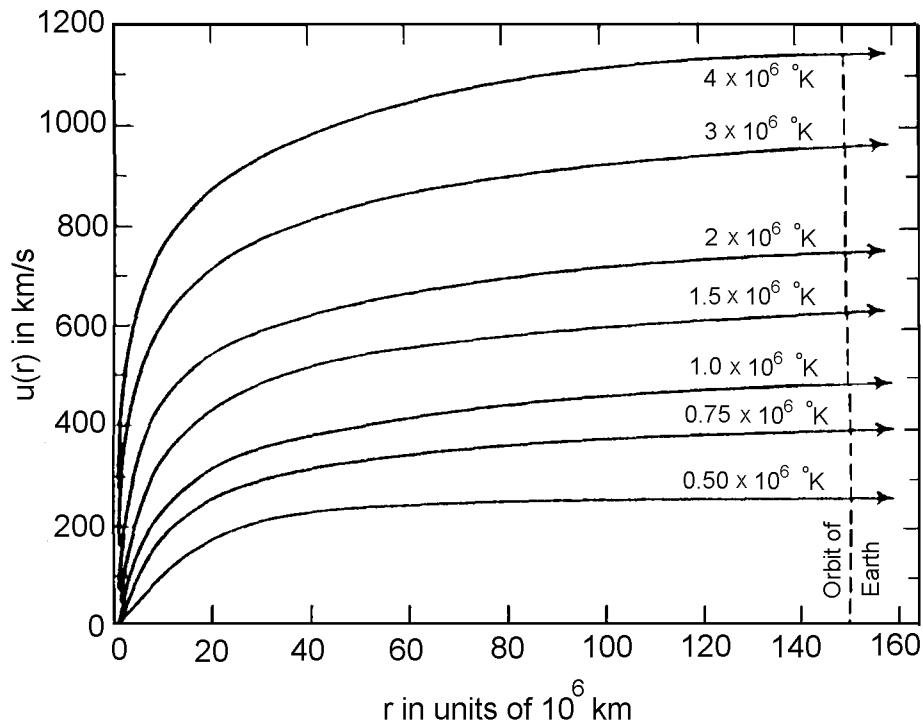


Figure 3.16 Expansion speed ($u(r)$) as a function of heliocentric distance for isothermal coronas of various temperatures.

Originally done in: E. N. Parker, Dynamics of the Interplanetary Gas and magnetic Field, *Astrophysical Journal*, 128, pages 664-676, 1958.

This version found in: A. J. Hundhausen, Coronal Expansion and Solar Wind, New York, Springer-Verlag, 1972, citing in turn Parker, E. N., Interplanetary Dynamical Processes, New York, Interscience Publishers, 1963.

This page intentionally left blank

Chapter 4 The Geomagnetic Field

A Introduction

During the past three decades observations have shown that magnetic fields exist throughout the observable universe. We have actually observed naturally occurring magnetic fields ranging from 10^{-10} T to 10^2 T and there are speculations that fields as high as 10^8 T may exist in neutron stars.

In addition to the observational data we also have a good start on what is believed to be a general theory for generation or rather amplification of large scale magnetic fields. The theory presumes the existence of large, electrically conducting objects which rotate and have internal energy sources. We are reasonably sure that both the geomagnetic and solar magnetic fields are maintained by this magneto hydrodynamic process.

Looking at the earth as an example of a planetary system, we recognize two force fields (a) the gravitational field without which there would be no earth (b) the magnetic field both within and around the earth. The most convincing way to see how our ideas about the geomagnetic field have changed in the last 30 years is shown in Figure 4.1 which shows excerpts from three textbooks published in 1951, 1960 and 1971 respectively.

1 Constituents of the Geomagnetic Field:

We usually consider the geomagnetic field to be composed of a Main Field whose shape is approximately dipolar and whose time variations are on the order of years or longer. To a first approximation we describe the field as being generated by a fictitious dipole near the center of the earth. Both the magnitude and the orientation of this dipole undergo slow time variations (secular variations) which give rise to the well known gradual changes of the geomagnetic field components.

Superimposed on this quasistatic, quasidipole field are small fluctuations having periods of days or less. The source of these fluctuations are generally exterior to the earth and while their amplitude is small ($\Delta B / B \lesssim 0.01$) they are of considerable interest both from a geophysical and an applications point of view.

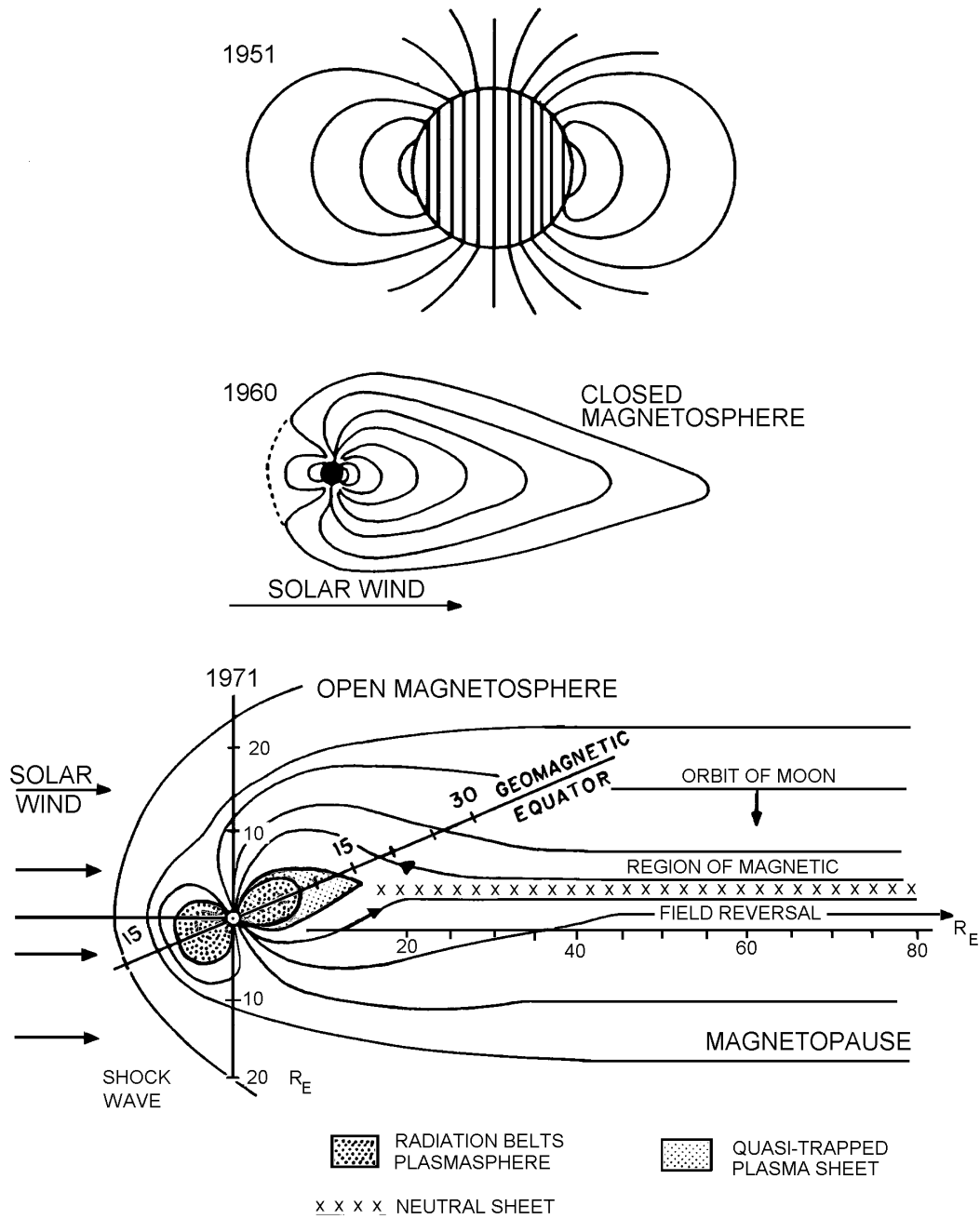


Figure 4.1

B Main Geomagnetic Field

1 The Dipole Field

The most elementary description of the earth's magnetic field at and above the surface is modeled after the field of a small current loop, i.e., a terrestrially centered magnetic dipole whose axis is tilted at an angle of approximately 11.5° with respect to the spin axis of the earth.

The sense of the field lines is from the southern geographic hemisphere toward the northern hemisphere as indicated in Figure 4.2. The magnitude and direction of the geomagnetic field vector at any point on the surface and at points above the surface out to a few earth radii can be obtained to an accuracy of about 10% using the following expressions for the field components of a dipole field in spherical coordinates.

$$\begin{aligned} B_r &= \frac{\mu_0}{4\pi} \frac{2m \cos \theta}{r^3} \\ B_\theta &= \frac{\mu_0}{4\pi} \frac{m \sin \theta}{r^3} \quad (\text{Eqn. 4.1}) \\ B_\phi &= 0 \end{aligned}$$

where:

$$\mu_0 = 4\pi \times 10^{-7} \text{ (Henry / meter)}$$

m = Magnetic Dipole moment of the earth
($8.1 \times 10^{22} \text{ m}^2\text{A}$)

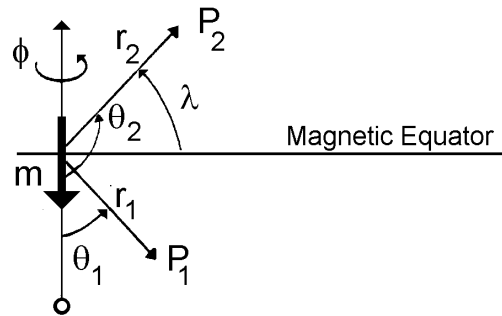


Figure 4.2. Coordinate system conventions for the earth's magnetic field.

It is often convenient to use the magnetic latitude λ (indicated in the sketch) rather than the polar angle θ . We note that in general

$$\theta = \lambda + \pi/2 \text{ where } \lambda > 0 \text{ in Northern Hemisphere}$$

$$\lambda < 0 \text{ in Southern Hemisphere}$$

Hence $\sin \theta = \cos \lambda$ and $\cos \theta = -\sin \lambda$, and we can also write the field components as

$$B_r = - \frac{\mu_0}{4\pi} \frac{2m \sin \lambda}{r^3} \quad (\text{Eqn. 4.2})$$

$$B_\lambda = \frac{\mu_0}{4\pi} \frac{m \cos \lambda}{r^3}$$

We can also calculate the magnitude of the \vec{B} vector:

$$B^2 = B_r^2 + B_\theta^2 = \left(\frac{\mu_0 m}{4\pi r^3} \right)^2 (4\cos^2\theta + \sin^2\theta)$$

or (Eqn. 4.3)

$$B = \frac{\mu_0 m}{4\pi r^3} \sqrt{1 + 3\cos^2\theta} = \frac{\mu_0 m}{4\pi r^3} \sqrt{1 + 3\sin^2\lambda}$$

Figure 4.3 shows the general shape of the dipole field. Note that the lines emerge from the earth in the Southern Hemisphere so that the pole located in Antarctica should really be called the geomagnetic North Pole. But it was felt that this confused everybody and it is called the GM South Pole.

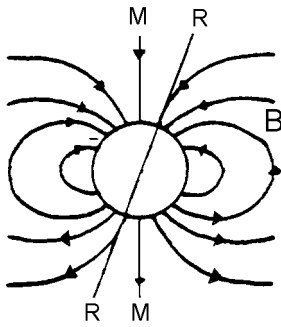


Figure 4.3 a) An idealized representation of the lines of **B** associated with the earth's magnetic axis and **RR** is its rotational axis.

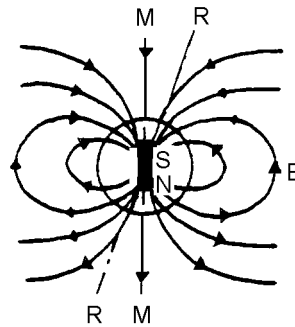


Figure 4.3 b) We can approximate the earth's internal magnetic field by imagining that a strong bar magnet is located at the center of the earth.

For computational purposes it is convenient to rewrite the equations for B_r , B_λ and B in units of earth radii and the field on the surface of the earth at the magnetic equator B_{os} .

To do this consider a point for which $\lambda = 0$ and $r = R_\oplus$.

Then $B_r = 0$ and $B_\lambda = \frac{\mu_0 m}{4\pi R_\oplus^3} = B_{os}$. We can use this relation to obtain:

$$\frac{\mu_0 m}{4\pi} = B_{os} R_\oplus^3 \quad (\text{Eqn. 4.4})$$

putting this constant back into the equations for B_r , B_λ and B we get:

$$\begin{aligned} B_r &= -B_{os} R_\oplus^3 \frac{2 \sin \lambda}{r^3} = -B_{os} \frac{2 \sin \lambda}{(r/R_\oplus)^3} \\ B_\lambda &= B_{os} R_\oplus^3 \frac{\cos \lambda}{r^3} = B_{os} \frac{\cos \lambda}{(r/R_\oplus)^3} \\ B &= B_{os} R_\oplus^3 \frac{\sqrt{1 + 3 \sin^2 \lambda}}{r^3} = \frac{B_{os} \sqrt{1 + 3 \sin^2 \lambda}}{(r/R_\oplus)^3} \end{aligned} \quad (\text{Eqn. 4.5})$$

These equations express radial distances in units of earth radii. The constant B_{os} has a value of $31.3\mu\text{T} = 3.13 \times 10^{-5}\text{T}$.

The points at which the axis of our fictitious dipole magnet intersect the surface of the earth are known as the geomagnetic poles. In 1980 the geomagnetic north pole was located at 78.8°N and 70.9°W .

There also exists something called the magnetic pole which is the place where the field is vertical. This is a place on the surface which must be determined experimentally and is difficult to find.

An improved approximation to the measured geomagnetic field is obtained by displacing the dipole from the planetary center. The best fit is obtained by a displacement of 436 km in the direction 15.6°N and 150.9°E which is toward the Pacific Ocean. This approximation describes the measured geomagnetic field to an accuracy of 2 to 3% and is known as the Eccentric Dipole Model.

2 International Geomagnetic Reference Field

The standard model for the Earth's magnetic field is the International Geomagnetic Reference Field (IGRF). This field is specified in terms of standard tables of values which can be inserted into a mathematical model termed a multipole expansion. A brief discussion of this process is presented here - application is tedious and best done with a computer.

On the surface of the earth and upward for approximately 50 km the atmosphere can be considered to be an electrical insulator and therefore there are no currents flowing in that region. If the current $i = 0$ then we can define a magnetic potential U from which the magnetic field \vec{B} can be derived

$$\vec{B} = -\nabla U.$$

Furthermore we can show that U must be a solution of Laplace's equation

$$\nabla^2 U = 0.$$

If we use spherical coordinates (r, θ, ϕ) with the origin at the center of the earth and the polar axis along the rotational axis of the earth we can show that the form of U is an expansion of the form

$$U(r, \theta, \phi) = \sum_{n=1}^{\infty} \sum_{m=0}^n \frac{r^n}{R_{\oplus}^{n+1}} (b_n^m \cos m\phi + c_n^m \sin m\phi) P_n^m(\cos \theta) + \sum_{n=1}^{\infty} \sum_{m=0}^n \frac{R_{\oplus}^{n+2}}{r^{n+1}} (g_n^m \cos m\phi + h_n^m \sin m\phi) P_n^m(\cos \theta) \quad (\text{Eqn. 4.6})$$

where R_{\oplus} is the radius of the earth and $P_n^m(\cos \theta)$ are the associated Legendre Polynomials (using Schmidt Normalization).

In the expression for U shown above, the first summation represents sources outside the earth (ionosphere and above) whereas the second summation represents sources internal to the earth. This is a powerful technique for separating these sources if we can determine the coefficients $b_n^m, c_n^m, g_n^m, h_n^m$ from actual measurements. Table 4.1 shows a recent set of coefficients derived from experimental measurements on the surface and magnetometers flown in low earth orbit.

Having obtained the proper expression for the potential U we still have to obtain the actual field components from $\vec{B} = -\nabla U$. If we define X , Y and Z as the north, east and vertically down components of \vec{B} we obtain

$$\begin{aligned} X = -B_\theta &= \frac{1}{r} \frac{\partial U}{\partial \theta} && \text{"North"} \\ Y = B_\phi &= -\frac{1}{r \sin \theta} \frac{\partial U}{\partial \phi} && \text{"East" (Eqn. 4.7)} \\ Z = -B_r &= \frac{\partial U}{\partial r} && \text{"Down"} \end{aligned}$$

where r increases outward, θ increases southward, and ϕ eastward in the usual arrangement for spherical coordinates.

The IGRF not only yields accurate (better than .5%) data for the field components at any point on or just above the surface of the earth, but it has also evaluated the contribution of external sources to the total field. We conclude that not more than a few tenths of a percent of the field intensity is due to sources external to the earth.

n	m	g	h	dg/dt	dh/dt	n	m	g	h	dg/dt	dh/dt
1	0	-29988		22.4		8	0	20		0.8	
1	1	-1957	5606	11.3	-15.9	8	1	7	7	-0.2	-0.1
						8	2	1	-18	-0.3	-0.7
2	0	-1997		-18.3		8	3	-11	4	0.3	0.0
2	1	3028	-2129	3.2	-12.7	8	4	-7	-22	-0.8	-0.8
2	2	1662	-199	7.0	-25.2	8	5	4	9	-0.2	0.2
						8	6	3	16	0.7	0.2
3	0	1279		0.0		8	7	7	-13	-0.3	-1.1
3	1	-2181	-335	-6.5	0.2	8	8	-1	-15	1.2	0.8
3	2	1251	271	-0.7	2.7						
3	3	833	-252	1.0	-7.9	9	0	6			
						9	1	11	-21		
4	0	938		-1.4		9	2	2	16		
4	1	783	212	-1.4	4.6	9	3	-12	9		
4	2	398	-257	-8.2	1.6	9	4	9	-5		
4	3	-419	53	-1.8	2.9	9	5	-3	-7		
4	4	199	-298	-5.0	0.4	9	6	-1	9		
						9	7	7	10		
5	0	-219		1.5		9	8	1	-6		
5	1	357	46	0.4	1.8	9	9	-5	2		
5	2	261	149	-0.8	-0.4						
5	3	-74	-150	-3.3	0.0	10	0	-3			
5	4	-162	-78	0.2	1.3	10	1	-4	1		
5	5	-48	92	1.4	2.1	10	2	2	1		
						10	3	-5	2		
6	0	49		0.4		10	4	-2	5		
6	1	65	-15	0.0	-0.5	10	5	5	-4		
6	2	42	93	3.4	-1.4	10	6	3	-1		
6	3	-192	71	0.8	0.0	10	7	1	-2		
6	4	4	-43	0.8	-1.6	10	8	2	4		
6	5	14	-2	0.3	0.5	10	9	3	-1		
6	6	-108	17	-0.1	0.0	10	10	0	-6		
7	0	70		-1.0							
7	1	-59	-83	-0.8	-0.4						
7	2	2	-28	0.4	0.4						
7	3	20	-5	0.5	0.2						
7	4	-13	16	1.6	1.4						
7	5	1	18	0.1	-0.5						
7	6	11	-23	0.1	-0.1						
7	7	-2	-10	0.0	1.1						

Table 4.1 Spherical harmonic coefficients of the IGRF 1980.0 in nT and nT/year for time derivatives. Handbook of Geophysics and the Space Environment, Air Force Geophysics Laboratory, 1985.

Field Components: There are several field components which can be used to specify the geomagnetic field vector \vec{B} at any point. In Figure 4.4 these components are illustrated. Any three will suffice to fix \vec{B} .

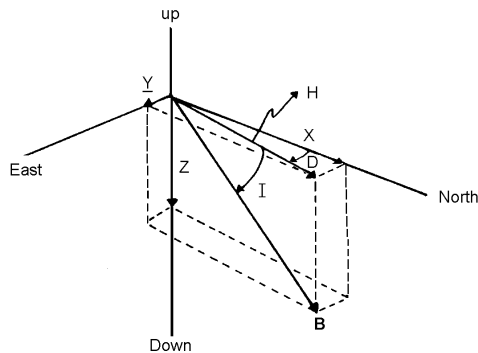
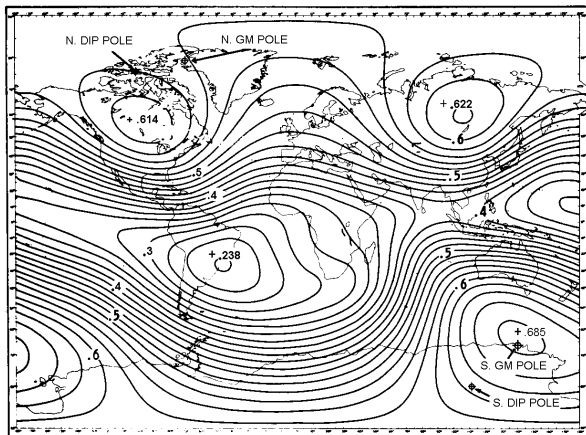


Figure 4.4

D = Declination (+ east from N)
 H = Horizontal component
 I = Dip Angle (+ down)
 B = total Intensity
 X = North-South component (+ North)
 Y = East-West component (+ East)
 Z = Vertical component (+ Down)

As an illustration of the results which can be obtained from such a model, Figure 4.5 shows the magnetic field intensity at the earth's surface. The contours are in Gauss. One Gauss is 10^{-4} Tesla



to the geomagnetic prime meridian in eastern hemisphere. Geomagnetic time can be defined in a manner analogous to geographic time.

Transformation Equations from Geographic to Geomagnetic Coordinates:

As mentioned earlier the location of the north geomagnetic pole as of 1980 is 78.8° N and 70.9° W ($= 289.1^\circ$ E). If we call the geographic latitude λ and the geographic longitude ϕ we can express the transformation equations as follows

$$\begin{aligned}\sin \lambda_m &= \cos \lambda_0 \cos \lambda \cos(\phi - \phi_0) + \sin \lambda_0 \sin \lambda \\ \cos \phi_m &= \frac{\sin \lambda_0 \sin \lambda \cos(\phi - \phi_0) - \cos \lambda_0 \sin \lambda}{\cos \lambda_m}\end{aligned}\quad (\text{Eqn. 4.8})$$

Where λ_m and ϕ_m are the geomagnetic latitude and longitude respectively and λ_0 and ϕ_0 are the geographic coordinates of the north geomagnetic pole.

Example:

Calculate λ_m and ϕ_m for Monterey whose geographic coordinates are 36.6° N and 121.9° W. ($= 238.1^\circ$ E)

$$\begin{aligned}\sin \lambda_m &= \cos(78.8^\circ) \cos(36.6^\circ) \cos(238.1^\circ - 289.1^\circ) + \sin(78.8^\circ) \sin(36.6^\circ) \\ &= 0.0981 + 0.5849\end{aligned}$$

$$\sin \lambda_m = 0.6830 \Rightarrow \lambda_m = 43.1^\circ$$

$$\cos \phi_m = \frac{\sin(78.8^\circ) \sin(36.6^\circ) \cos(-51.0^\circ) - \cos(78.8^\circ) \sin(36.6^\circ)}{\cos(43.1^\circ)}$$

$$= \frac{0.4956 - 0.1158}{0.7304} = 0.52$$

$$\phi_m = 58.7^\circ \text{ or } 360^\circ - 58.7^\circ = 301.3^\circ$$

Thus the geomagnetic coordinates of Monterey are 43.1° N, 301.3° E.

4 Dynamo Theory of the Main Field

During the past 30 years the outlines of a general theory capable of explaining the essential features of the Main Field have emerged. Furthermore it is believed that this general approach will apply to any astronomical body for which the following three conditions are met

- (1) The system must be large and have a high electrical conductivity.
- (2) There must be some non-symmetric rotational or convective motion present.
- (3) There must be an internal energy source available.

The liquid core of the earth appears to meet these criteria

- (1) The outer core extends from a depth 2800 km to a depth of 5400 km and consists of liquid metals (92% Fe and about 8% Ni) and is believed to have an electrical conductivity of about 10^6 (Siemens).
- (2) As a result of viscous forces and temperature gradients in the molten core there is both differential rotation and turbulent upwelling present. The secular changes in the geomagnetic field seem to confirm these assumptions.
- (3) Several energy sources are potentially available in the core and specially in the inner core (below 5400 km). These are the energy generated by radioactive decay, by gravitational contraction and possibly energy released by high pressure phase transitions.

The details of the magneto hydrodynamic theory are still quite uncertain as well as very complex. We shall therefore give only a qualitative description of the major points of the theory.

We may visualize the generation of the main geomagnetic field something like this: A field line of external origin (perhaps due to the solar or galactic field is "frozen" into the highly conducting core of the earth and due to rotation of the core is wound up into a strong (.05T or 50 Gauss) azimuthal field as illustrated in Chapter 1. By a series of complex motions associated with turbulent upwelling this azimuthal field is carried outward and converted into a toroidal helical field and an associated azimuthal current system which in turn generates the observed dipole field of the earth .

Although many aspects of this theory are still quite speculative, considerable support for the general concept comes from solar observations. On the sun there is no solid mantle and hence the azimuthal field is directly observable on the surface of the sun. This "dynamo theory" is able to explain at least in principle the geologically documented reversals of the geomagnetic field. These reversals are believed to be caused by the interactions of two or more "dynamoes" operating within the core.

Finally it should be noted that the theory actually describes a "flux amplification" process in which energy from an internal source is used to amplify a weak seed field into a much stronger dipole field. This mechanism is believed to be active in all astronomical bodies (planets, stars, galaxies) which satisfy the original conditions stated above. The ultimate source of the seed field must of course be an electric current and a number of processes are available to generate such initial currents.

C Time Variations of the Geomagnetic Field

The Main Field which we discussed so far shows very slow time variations both in magnitude & direction, characteristic times being on the order of thousands of years. These so-called "secular variations" are of geologic origin and result among other things in the well known variations of magnetic north with respect to geographic north (variation of declination)

1 Diurnal Variations

When continuous records from magnetic observatories are examined the three magnetic field components show variations which are repeated every day, mostly during the daylight hours. The pattern of each vector variation also changes systematically with latitude on a global scale, as shown in Figure 4.6.

Analysis shows that about two thirds of the variations are due to sources external to the earth and one third appears to be due to internal sources. However the internal sources are the result of electric currents induced within the earth by external time varying magnetic fields. Thus all of the variations are directly or indirectly due to external sources.

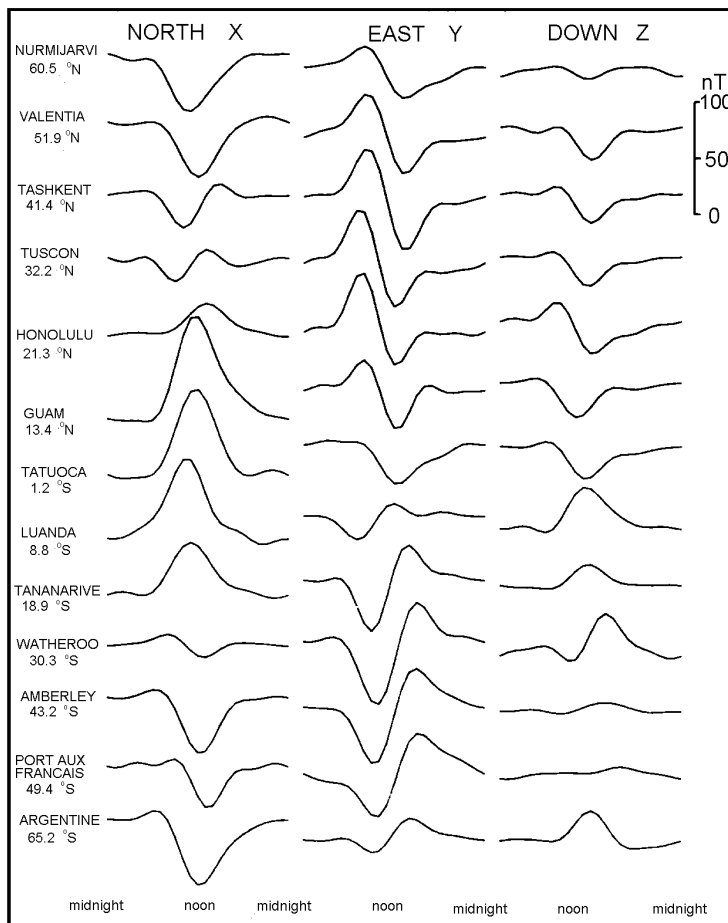


Figure 4.6 Diurnal variation curves as a function of local solar time for observatories at various (geographic) latitudes. From: Introduction to Geomagnetism, W. D. Parkinson, page 262.

Ionospheric current systems are responsible for the daily variations. Convective movements of the conducting upper atmosphere across the earth's magnetic field produce electric currents in the medium (ionosphere). The large scale motions of the upper atmosphere occur because of pressure and temperature differences brought about by solar heating, tidal forces and the Coriolis force. This complex mechanism is usually referred to as the atmospheric dynamo. An idealized current system for the generation of the observed daily variations is shown in Figure 4.7

Strong currents in the system are mostly found on the sunlit side of the earth and in the region between the equator and mid latitudes. In each hemisphere there is a vortex of current with its center at about 30° geomagnetic latitude and near the noon meridian. The current rotates counterclockwise in the northern hemisphere and clockwise in the southern hemisphere. Although the current densities are quite low the total circulating current is about 120,000 A. This current system is fixed relative to the sun-earth line and as a given point on the surface rotates under it the magnetic field of the overhead currents gives the variations shown in Figure 4.6.

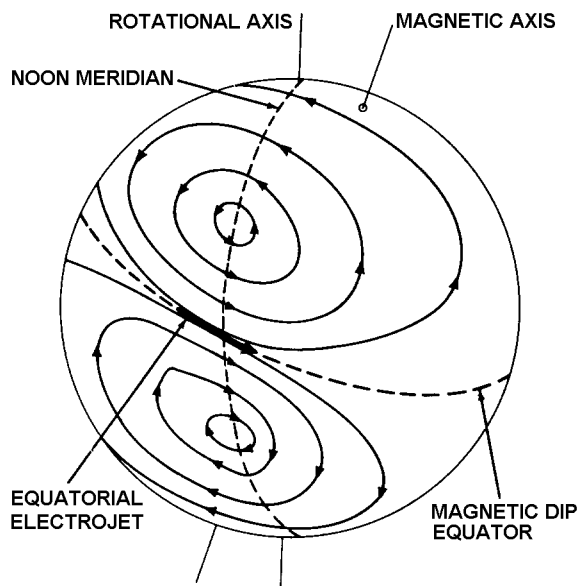


Figure 4.7 Currents flow in the upper atmosphere (ionosphere) as shown schematically, in the directions indicated by the arrows.

2 Magnetic Storms

Intense, worldwide transient variations of the geomagnetic field are called magnetic storms. They are frequently accompanied by other phenomena such as intense auroral displays, ionospheric disturbances, etc.

The frequency of magnetic storms correlates closely with the solar sunspot cycle leading to the conclusion that magnetic storms on earth are caused by solar events. The cause of magnetic storms is the arrival of a solar plasma pulse ejected from the sun during a solar flare. The plasma traveling at ~ 1000 km/sec reaches the earth a day or two after the occurrence of the flare. What we describe as a magnetic storms is the result of the interaction of the solar plasma with the earth's magnetosphere. Most magnetic storms undergo a pattern of development such as that indicated in Figure 4.8.

The start of a typical storm is marked by an abrupt increase in the horizontal component (H) known as the sudden commencement (SC) of the storm. The increase is typically 30-50 nT world-wide in a matter of a few minutes. The effect is the result of the impact of the plasma pulse traveling at several times the ordinary solar wind velocity on the outer boundary of the magnetosphere. The compression of the boundary is transmitted in the form of magneto hydrodynamic waves and is seen on the surface of the earth as an increase in the ambient field strength.

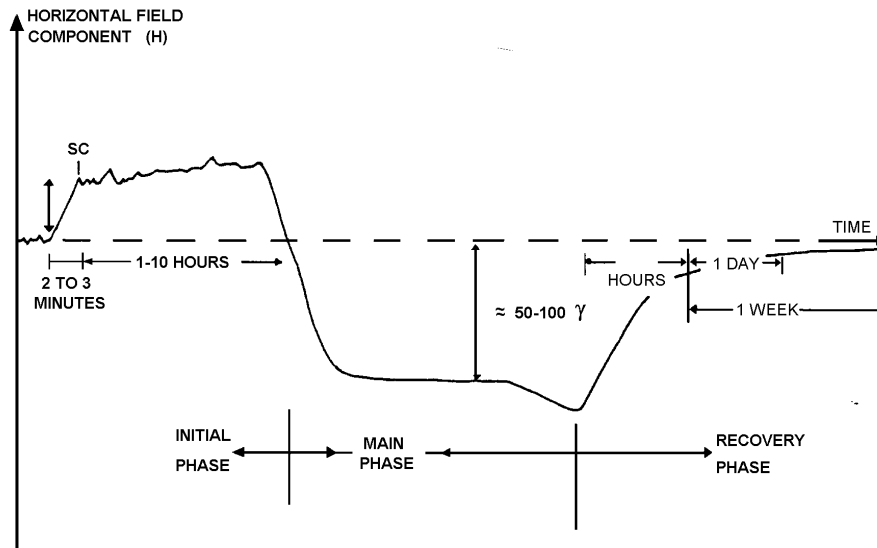


Figure 4.8 Example of magnetic field measurements during a magnetic storm. The baseline (horizontal axis) represents mean magnetospheric conditions. The symbol Sc is used to denote the start of the sudden commencement which typically has an intensity of 5 - 30 gammas (γ) (or nano-Tesla, nT). Prochaska, 1980).

For the next 2-10 hrs. the horizontal field component remains above its undisturbed value. This interval is known as the initial phase of the storm and corresponds to a compressed state of the magnetosphere. This new equilibrium state persists until the bulk of the plasma has passed the earth on its outward flow through the solar system.

Meanwhile the main phase of the storm lasting 12 to 24 hrs. is initiated: H decreases to values typically 100 nT below prestorm values. This decrease is the result of a westward ring current induced in the magnetosphere by the compression during the initial phase. The ring current is composed of newly trapped particles (both electrons & protons) as well as acceleration of particles previously present in the magnetosphere at altitudes of several earth radii. During the final stage of the storm lasting several days the ring currents gradually dissipate and the field returns to its prestorm value.

Individual storm records show great irregularities as compared to the "typical" pattern of Figure 4.8. The initial and main phase tend to be very noisy and large amplitude fluctuations occur with periods as long as 30 minutes. Storm records in the auroral and polar regions are particularly noisy as indicated in Figure 4.10 which shows a storm record at 4 magnetic latitudes. Note also the difference in the scales on left side of the graph.

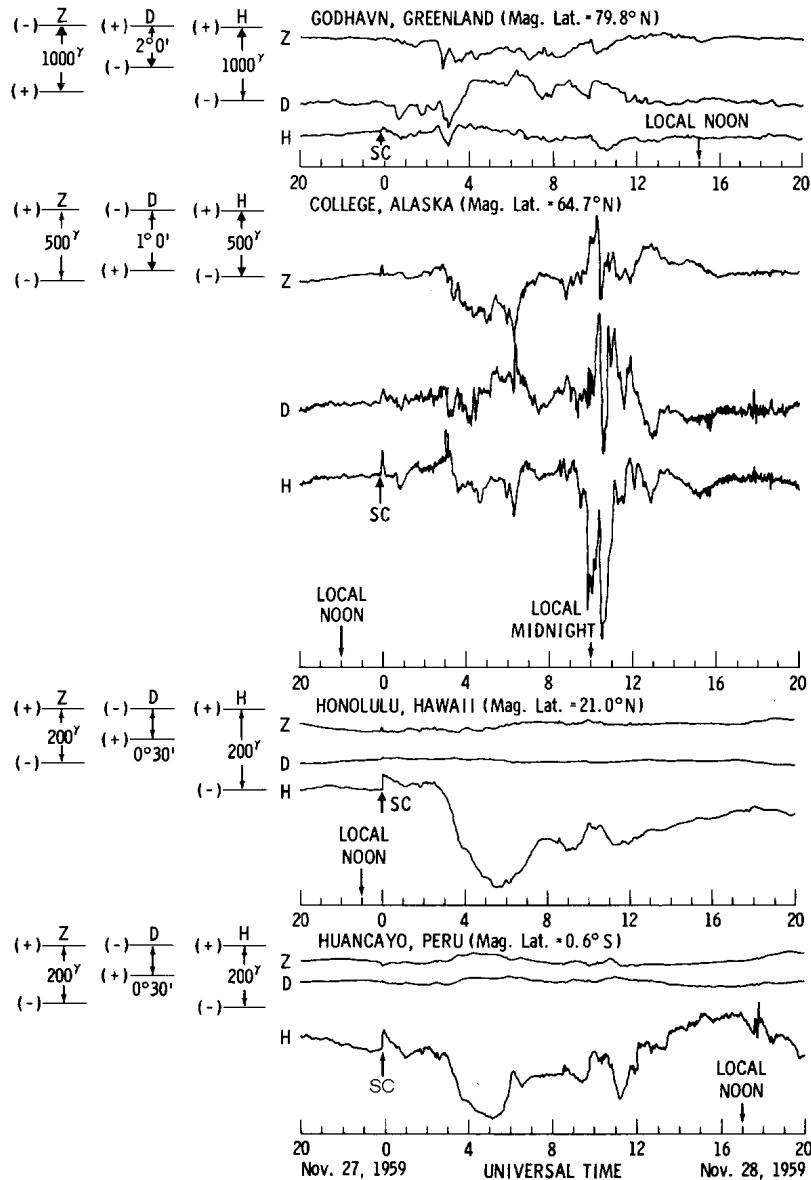


Figure 4.9 Examples of magnetic storm of moderate intensity as recorded at different latitudes. H, D, and Z are the symbols for the three components of the magnetic field, Horizontal, Declination, and Vertical, respectively.

From: Introduction to Space Science, edited by Wilmot Hess, Chapter 1, the Earth's Magnetic Field, by M Sugiura and J. P. Heppner, page 59, 1965.

Not all solar flares (or CME's) produce magnetic storms on earth. The actual trajectory which a given plasma pulse follows after leaving the sun is generally quite complex. The question of whether or not it will reach the earth depends upon where it leaves the sun, and then more subtle problems involving the transport of plasma. The resulting transport pattern is sufficiently complex that it is not presently possible to predict whether the plasma stream from a given flare will hit or miss the earth.

3 Magnetic Indices

A large number of indices have been devised to characterize the state of activity of the geomagnetic field or some portion of it. Most of these indices are highly specialized and will not be discussed here. However two of them, the K and Dst indices, have general use and we shall have a look at them:

The K index is an indicator of the general level of activity in the magnetic field, at mid to high latitudes. As such, it is often used as an indicator of auroral activity, and the level of magnetic activity at geosynchronous orbit.. Each observatory assigns a digit between 0 and 9 for each 3 hr. interval starting at 0^h (UT) to each of the three field components (X, Y, Z or H, D, Z usually). The largest deviation from the average is used and normalized on a quasi-logarithmic scale. This means that each observatory chooses an appropriate scale so that the frequency of a given K number is the same at all observatories. Thus 350 nT may correspond to K = 9 at one observatory but for another station K = 9 corresponds to 1000 nT.

Each K value represents roughly a factor of 2 increase in the magnitude of the excursion. It has also become customary to subdivide each value on the K scale in to 3 sub-steps by the subscripts -, 0 and +. Thus we have

---- 4₋, 4₀, 4₊, 5₋, 5₀ ----

The planetary index K_p is obtained by averaging the standardized K values from 12 observatories located between 48° and 63° geomagnetic latitude. Figure 4.10 shows a typical magnetogram, and the K value which would correspond to each level of magnetic field fluctuation. The K value for the Fredricksberg, VA observatory is often used as a "local" index for the continental US. It is important to note that since we are averaging a set of observations taken three hours apart we cannot get any information about higher frequency oscillations i.e., for $f > 1/(3 \times 3600) \sim 10^{-4}$ Hz

Diagrams of K_p are typically made for long periods as illustrated in Figure 4.11. Note that the horizontal axis is 27 days - one solar rotation. Can you see the periodicity?

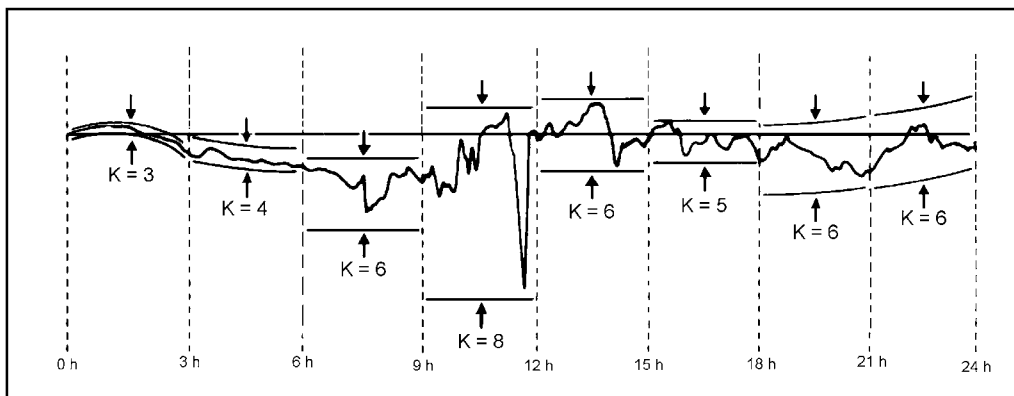


Figure 4.10 A typical magnetogram (H trace only) showing the ranges corresponding to the K-indices assigned to each 3-hour interval for that day.

From: Introduction to Geomagnetism, W. D. Parkinson, page 284.

PLANETARY 3-HOUR-RANGE INDICES (Kp) BY 27-DAY SOLAR ROTATION INTERVAL

University of Gottingen

Kp through December 31, 1993

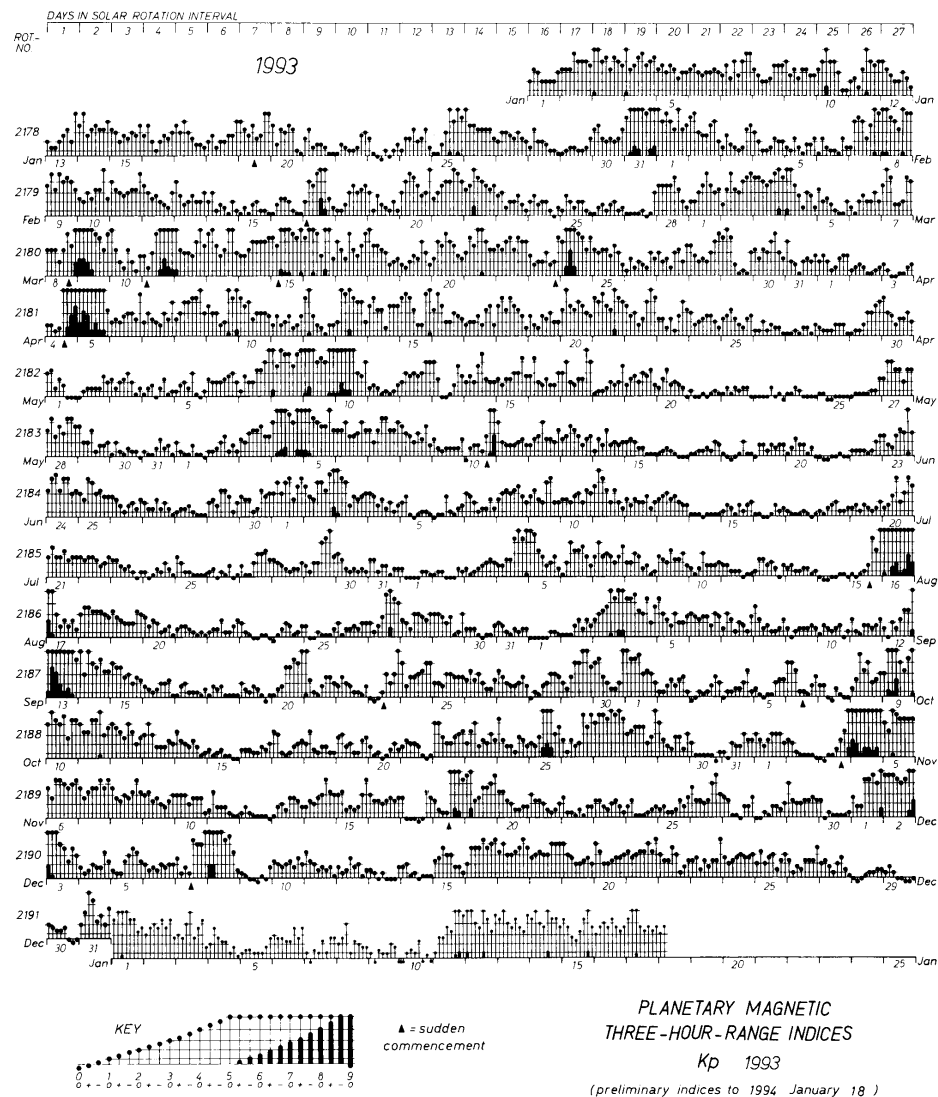


Figure 4.11 Kp figures for one year

A practical measure for the overall strength of the extraterrestrial ring current is the D_{ST} -index which measures the middle latitude spatially averaged decrease in the horizontal component H of the earth's surface magnetic field $D_{ST} = \langle D_H \rangle$. Under this definition the quiet time ring current corresponds to $D_{ST} = 0$. Hourly values of D_{ST} index are published by NASA National Space Science Data Center, Goddard Space Flight Center, Maryland. Magnetic storms generally have D_{ST} depressions on the order of 100 to 200 nT (very large storms may exceed $D_{ST} = 300$ nT), and the D_{ST} index may also fluctuate substantially for other geomagnetic conditions for which D_{ST} generally remains less than 50 nT. Figure 4.12 shows an example of the D_{ST} index plotted for June - December 1972, and the occurrence of four magnetic storm periods in June, August, September and October/November is evident. Note the corresponding increase in auroral activity, as revealed by the periods of high $K_p > 5$.

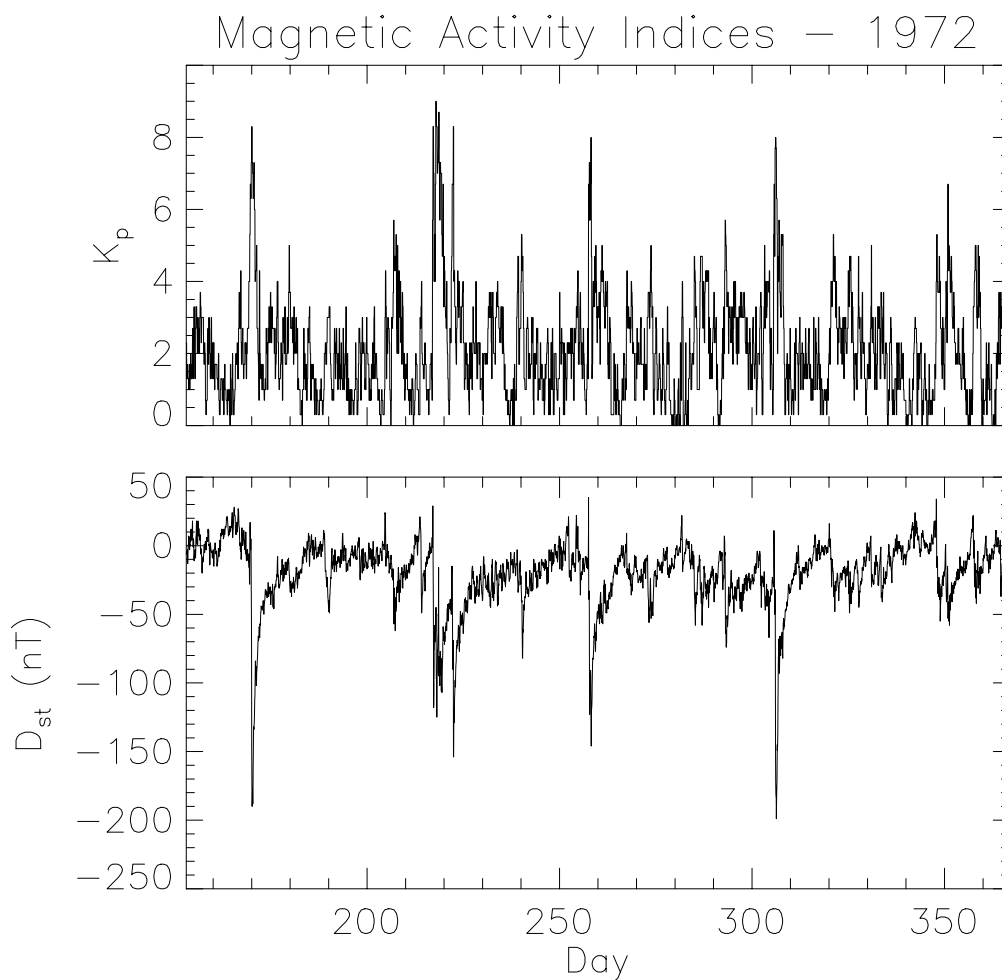


Figure 4.12 Magnetic Activity Indices

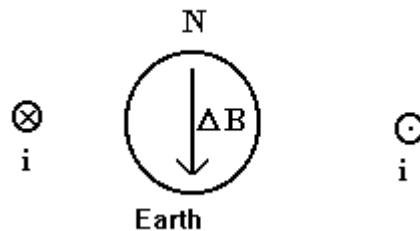
D References

Nishida, A. (1978). Geomagnetic Diagnosis of the Magnetosphere. Springer-Verlag, N. Y.

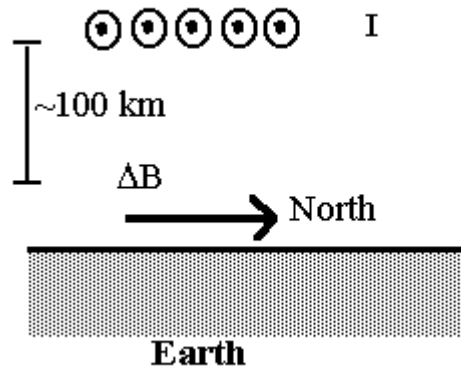
Parkinson, W. D., Introduction to Geomagnetism, Elsevier Science Pub Co., Inc., NY, NY

E Problems

1. The activity index, Dst, can reach ~ 100 nT. Estimate the current necessary to produce this magnetic field effect. Assume the current is a "ring" around the earth, at $3 R_E$. That is, if $\Delta B \sim 100$ nT, use the Biot-Savart law for a loop to estimate the current, i . To keep life simple, assume that the standard formula (e.g. Halliday and Resnick) for the field at the center of a ring is reasonably close to the formula for the field off the axis (e.g. the surface of the earth).



2. The diurnal current variations are due to currents flowing in the upper atmosphere (ionosphere). Take this current to be an infinite, planar (e.g. sheet) current. How large a current density is needed to produce a 100 nT perturbation (ΔB)?



3. Find the magnetic latitude and longitude of Diego Garcia. Find both components of the local magnetic field there.

Chapter 5 The Magnetosphere

A Introduction

In the absence of an interplanetary plasma, the earth's dipole field would extend indefinitely in all directions. However, the geomagnetic field produces a semipermeable obstacle to the solar wind and the resulting interaction produces a cavity around which most of the plasma flows. The extent of this obstacle, called the magnetosphere, is related to the solar wind density and velocity and the interplanetary magnetic field (IMF). This cavity is filled by plasma which largely originates in the earth's upper atmosphere.

We may thus describe the magnetosphere (Figures 4.1 and 5.1) as the region of space surrounding the earth in which the geomagnetic field plays a dominant role. Before describing this region we need to do some more basic physics.

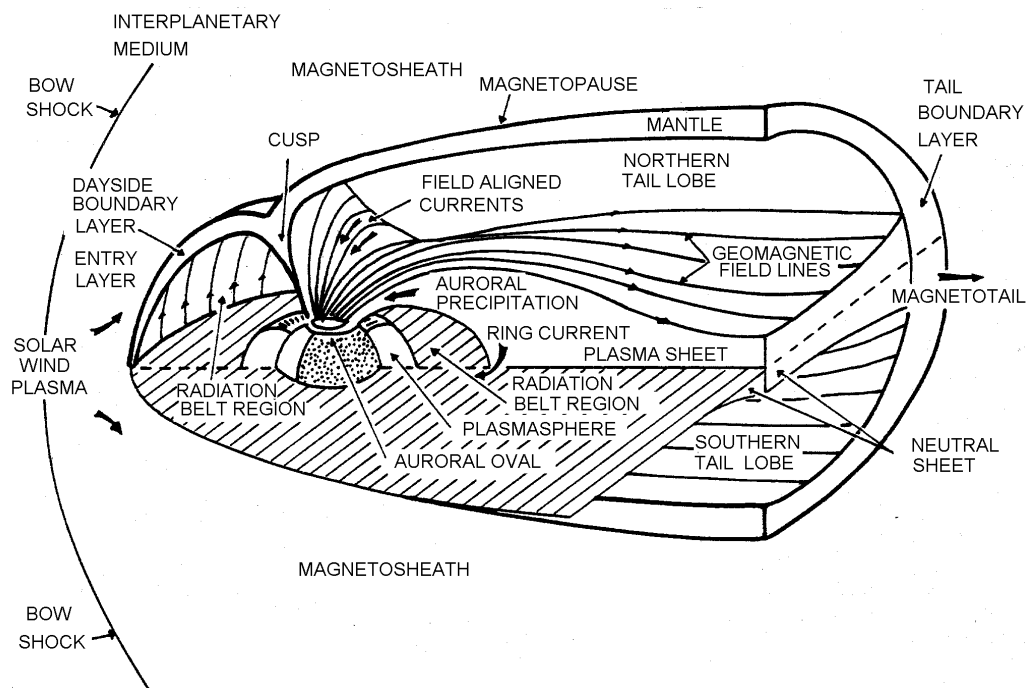


Figure 5.1 The Magnetosphere

B Plasma Physics

1 The Magnetic Moment and Mirroring

Let us begin by defining two quantities which are needed to develop the concept of magnetic mirroring. These are the magnetic flux through a surface, and the magnetic moment. The magnetic moment is one of the most useful concepts in magnetospheric physics, indeed all of plasma physics. It is almost always a conserved quantity, and hence enormously useful in determining the behavior of charged particles.

The magnetic moment of a current loop is defined (see figure 5.2) as a vector, $\vec{\mu}$, given by

$$\vec{\mu} = i \vec{A} \quad (\text{Eqn. 5.1})$$

Here, i is the circulating current, and \mathbf{A} is the area vector, defined as positive outward from a closed surface. For a flat surface we use the right hand rule to define the positive direction. In order to make this quantity useful, we need to relate it to two important characteristics of the plasma - the kinetic energy of the charged particles, and the strength of the ambient magnetic field.

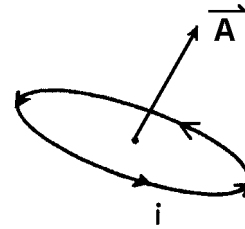


Figure 5.2

Consider now the case where the current i consists of a single circulating particle of charge q , in an external magnetic field, \mathbf{B} . This circulating charge represents a current:

$$i = qf_c = q\omega_c/2\pi$$

where ω_c is the cyclotron- or gyro- frequency (in radians/second) in an external magnetic field. The magnetic moment of the orbit is:

$$\mu = iA = (q\omega_c/2\pi)(\pi r_c^2) \quad (\text{Eqn. 5.2})$$

We have from before:

$$\omega_c = qB/m \quad \text{and} \quad r_c = mv_{\perp}/qB$$

where v_{\perp} is the component of the velocity vector perpendicular to \vec{B} .

Plugging these into equation 5.2 gives us:

$$\mu = \left(\frac{q^2 B}{2m} \right) \left(\frac{m^2 v_{\perp}^2}{q^2 B^2} \right) = \frac{\frac{1}{2} m v_{\perp}^2}{B}$$

or

$$\mu = \frac{K_{\perp}}{B}$$

(Eqn 5.3)

Here, K_{\perp} represents the part of the particles kinetic energy which is associated with the velocity component perpendicular to B . The division of the velocity vector between field-aligned and perpendicular components requires the definition of a new concept - the "pitch angle". Figure 5.3 illustrates the concept.

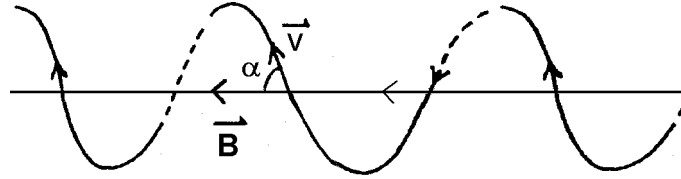


Figure 5.3 - Illustration of pitch angle definition (α). Note that the pitch of a screw is the same concept.

We see from figure 5.3 that:

$$v_{\perp} = v \sin(\alpha) \quad (\text{Eqn. 5.4})$$

and hence the magnetic moment (μ) is:

$$\mu = \frac{\frac{1}{2}mv^2 \sin^2(\alpha)}{B} \quad (\text{Eqn. 5.5})$$

Why is this quantity conserved? The answer lies in Faraday's Law (and Lenz's Law; Halliday and Resnick, Fundamentals of Physics, chapter 32 of the third edition)

Recall that the magnetic flux, ϕ , through a given surface is defined (see figure 5.2) as:

$$\phi = \int_S \vec{B} \cdot d\vec{S} \quad \text{Eqn 5.6}$$

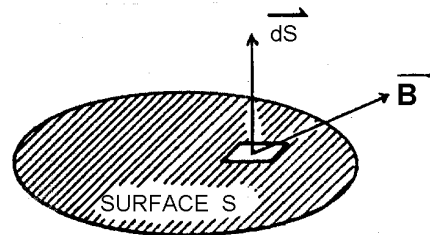


Figure 5.4

This quantity is, in some sense, the "amount" of field which crosses the surface defined by S . Its units are Tesla-meter-squared (Tm^2), or the Weber (W). What is the relationship between this quantity and the magnetic moment? It turns out that they are the same, within a (constant) fudge factor.

$$\phi = B (\pi r_c^2) \quad (\text{Eqn. 5.6})$$

(for reasonably constant B)

$$= \pi B \left(\frac{m^2 v_{\perp}^2}{q^2 B^2} \right) = \frac{2\pi m}{q^2} \left(\frac{mv_{\perp}^2}{2B} \right) \quad (\text{Eqn. 5.7})$$

or

$$\phi = \frac{2\pi m}{q^2} \mu \quad (\text{Eqn. 5.8})$$

Faraday's Law says that a voltage, or emf, \mathcal{E} , will be induced around the orbit if Φ changes in time.

$$\mathcal{E} = \frac{-d\Phi}{dt} \quad (\text{Eqn. 5.9})$$

(Lenz's law then applies this to indicate that attempts to change the magnetic flux through a loop will induce a current in just such a way as to oppose the change in flux.) How does this apply in a plasma? Consider the case where the particle is moving (typically along the magnetic field), and the amplitude of the magnetic field is changing. What happens to the magnetic flux?

A first look at Faraday's law suggests a problem. The emf produced by the changing magnetic flux would result in an acceleration of our charge and hence a change in its kinetic energy. We know, however, that the total kinetic energy of our particle does not change (static magnetic fields can not do work!), which implies that $\Phi = \text{a constant}$ (through the orbit). How do we resolve this contradiction? Nature resolves this apparent contradiction by forcing the pitch angle to change.

Therefore

$$\mu = \frac{\frac{1}{2}mv^2 \sin^2(\alpha)}{B} = \text{a constant} \quad (\text{Eqn. 5.10})$$

What is the net effect of this? As the particle penetrates into regions of larger magnetic field (B), the pitch angle (α) must increase in order to keep the magnetic moment, μ , constant. Sooner or later α will equal 90° and the longitudinal motion stops. Careful analysis shows that there is a small, non-zero force in a converging magnetic field geometry which produces this effect.

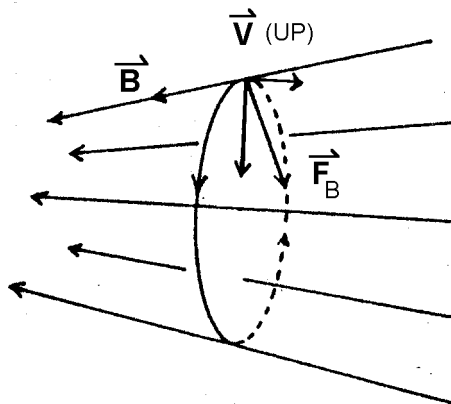


Figure 5.5

The magnetic force, F_B , will have a component in the plane of the orbit and a longitudinal component pointing into the region of lower magnetic field (to the right, here).

If the pitch angle is 90° at the point where the magnetic field is B_M , then

$$\frac{\frac{1}{2}mv^2 \sin^2(\alpha)}{B} = \frac{\frac{1}{2}mv^2 \sin^2(90^\circ)}{B_M} \Rightarrow \sin^2(\alpha) = B/B_M \quad (\text{Eqn. 5.11})$$

where α is the pitch angle at the position where the field is B , and B_M is the value of the field where the pitch angle becomes 90° and the particle is reflected (i.e. mirrors).

Example: the initial pitch angle of a particle at the magnetic equator is 30° , so $\sin^2 \alpha = 0.25$, hence the ratio of the magnetic field at the mirror point to that at the equator is 4:1. If the equatorial magnetic field had a value of 100 nT (as at geosynchronous orbit), it would mirror at a latitude where $B = 400$ nT.

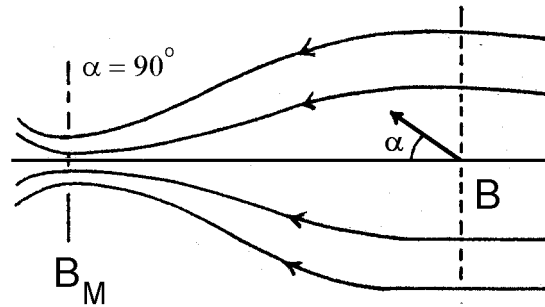


Figure 5.6 - magnetic field lines converge to the left

The concept of the loss cone can now be developed. This idea is important for understanding how particles are lost in the magnetosphere.

If we rewrite the mirror condition, found above, as:
$$\frac{\sin^2(\alpha)}{B} = \frac{\sin^2(\alpha_M)}{B_M}$$

we can see that particles whose pitch angle at a magnetic field strength, B , is less than α will have $\alpha_m < 90^\circ$, and hence will not be reflected and can pass through the mirror region. This is generally the means by which particles escape such magnetic field geometries. Typically, satellite measurements are made at or near the magnetic equator, and this is the minimum B region in the above model. The mirror point, then is at the top of the atmosphere, at high latitude, where the field is much stronger.

Example:

Start at the magnetic equator, at an altitude of one earth radius. Take the magnetic field strength to be 3.75×10^{-6} Tesla, and look for a mirror altitude near the surface of the earth, $B = 4.5 \times 10^{-5}$ T. Mirroring will occur at or before that point if α meets the condition:

$$\frac{\sin^2(\alpha)}{3.75 \times 10^{-6}} = \frac{\sin^2(90)}{4.5 \times 10^{-5}} \text{ or } \sin^2(\alpha) = \frac{3.75 \times 10^{-6}}{4.5 \times 10^{-5}} = 0.083$$

giving $\alpha = 16.8^\circ$. Particles with equatorial pitch angles less than this value will not mirror, and hence will hit the surface of the earth.

We thus see that the charged particle which in general will execute a spiral motion around a given field line will move from the equator into a region of higher magnetic field and that we thus have the likelihood of mirroring occurring, as illustrated in figures 5.6 and 5.7.

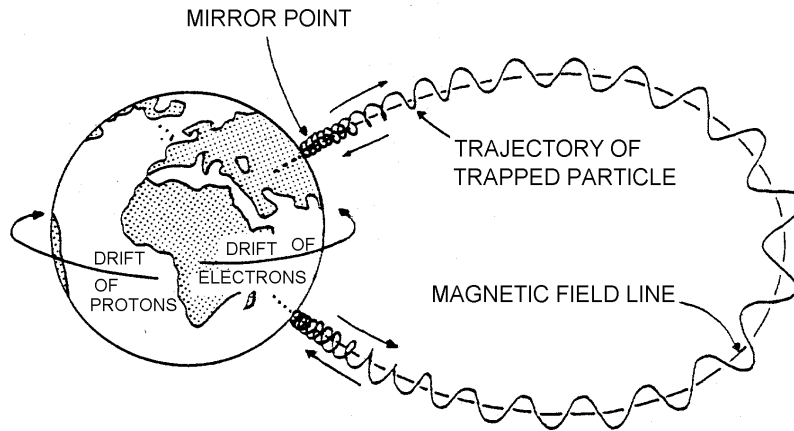


Figure 5.7 Motion of charged particles trapped by the earth's magnetic field. The mirror point is the position where a charged particle stops and changes direction (i.e. a trapped particle bounces between its mirror points). Besides the two-degree motion of spiraling back-and-forth along the field line, the charged particles also have a third degree of motion, a drift around the earth. Notice that electrons drift eastward and protons drift westward, and the resulting charge separation produces a ring current system (After Stern and Ness, 1981).

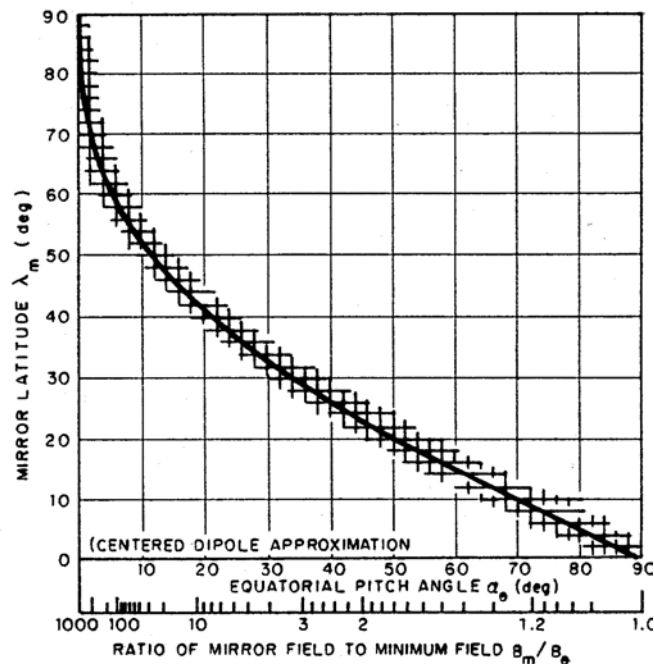
Consider a particle which has a pitch angle α_0 when it crosses the equatorial plane where the field is B_0 . We saw above that the particle will not mirror if the angle α_0 is less than α_c given by:

$$\sin^2(\alpha_c) = B_0 / B_M \quad (\text{Eqn. 5.12})$$

where B_M = Mirror field strength

Figure 5.8 Mirror latitudes in a dipole magnetic field

Figure 5.8 presents the relationship between the equatorial pitch angle, α_0 , the required ratio of the mirror field B_M to equatorial field B_0 and the latitude at which the particle will mirror. If we decrease the equatorial pitch angle α_0 , the required magnetic field ratio, B_M/B_0 , increases. This means that the particles will mirror at higher latitudes. Since there is a maximum value of B_M which can be reached along any magnetic field line (before hitting the atmosphere), particles with initial pitch angles which are small enough will not be mirrored.



2 Drifts

In addition to the gyration and bounce motions, the entire plane of the particle trajectory will precess in the azimuthal direction around the earth. This component of the motion is essentially a drift motion caused by the fact that the field is not uniform. The dipole magnetic field has both a gradient and curvature at each point. These cause drifts, much as the external electric or gravitational fields were shown to cause drifts.

We can show that the net force on a particle of magnetic moment μ in a non-homogeneous magnetic field is given by:

$$\vec{F} = -\mu \nabla |\vec{B}| \quad (\text{Eqn. 5.13})$$

Since in general the drift velocity is given by:

$$\vec{v}_D = \frac{1}{q} \frac{\vec{F} \times \vec{B}}{B^2} \quad (\text{Eqn. 5.14})$$

we get for the gradient drift (or so-called grad-B drift):

$$\vec{v}_D = \frac{-\mu}{q} \frac{\nabla |\vec{B}| \times \vec{B}}{B^2} = \frac{\mu}{q} \frac{\vec{B} \times \nabla |\vec{B}|}{B^2} \quad (\text{Eqn. 5.15})$$

This can also be expressed in terms of the transverse (or perpendicular) kinetic energy, since we had from the definition of the magnetic moment:

$$\mu = \frac{K_{\perp}}{B} = \frac{\frac{1}{2} m v_{\perp}^2}{B}$$

the drift velocity can therefore be expressed:

$$\vec{v}_D = \frac{K_{\perp}}{qB} \frac{\vec{B} \times \nabla B}{B^2} \quad (\text{Eqn. 5.16})$$

This drift is associated with variations in the strength of the magnetic field, in directions perpendicular to B . For a dipole field, this is primarily the variation in magnitude with radius. There is also a term associated with the shape of the field, termed the curvature drift. It can be thought of as a result of the centrifugal force exerted on a particle moving along a curved magnetic field line. The magnitude of this force is:

$$F_c = \frac{mv_{\parallel}^2}{R} \quad \text{Eqn. 5.17}$$

Here, v_{\parallel} is the velocity component of the particle along the magnetic field line ($v \cos \alpha$), and R is the radius of curvature of the field line along which the guiding center is moving. This results in the curvature drift given by:

$$\vec{v}_c = \frac{mv_{\parallel}^2}{qR_c^2} \frac{\vec{R}_c \times \vec{B}}{B^2} \quad (\text{Eqn. 5.18})$$

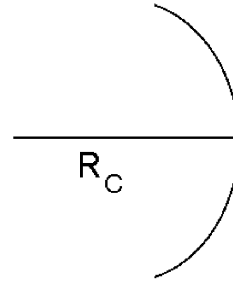


Figure 5.9 Magnetic field lines curve with radius R

The curvature and gradient of a magnetic field are intimately related, due to the need to satisfy Maxwell's equations. This is particularly so in a dipole field. We can combine the terms for the gradient and curvature drift to give:

$$\begin{aligned} \vec{V}_d + \vec{V}_c &= \frac{m}{q} \frac{\vec{R}_c \times \vec{B}}{R_c^2 B^2} (v_{\parallel}^2 + \frac{1}{2} v_{\perp}^2) \\ &= \frac{m}{q} \frac{\vec{B} \times \nabla B}{B^3} (v_{\parallel}^2 + \frac{1}{2} v_{\perp}^2) \end{aligned} \quad (\text{Eqn. 5.19})$$

where we have made use of the identity:

$$\frac{\nabla B}{B} = -\frac{\vec{R}_c}{R_c^2}$$

The precession of the "bounce plane" is due to the drifts given by Eqn. 5.19 above with the directions of precession for positive and negative charges as shown in Figure 5.7. Positive charges drift west, negative charges drift east. Both signs contribute to a current which is westward.

The totality of these gyrating, bouncing and precessing protons (and other ions) and electrons constitute the radiation belts, often referred to as the Van Allen Belts, after their discoverer (and the students who did the work....)

3 Characteristic Time Scales

The corresponding time scales for a 1 MeV particle at an altitude of 2000 km are:

	Electrons	Protons
Gyration Period	7×10^{-6} sec	4×10^{-3} sec
Bounce Period	0.1 sec	2.2 sec
Drift Period	53 minutes	32 minutes

(Differences here are due to the fact that at these energies, the electrons are highly relativistic, and all the relativistic terms we've ignored come into play.)

The substantial difference in characteristic time scales for each of the three characteristic motions makes it possible to consider the motions as decoupled from each other. This enables a great simplification of our studies.

Example:

As we saw in Chapter 1 the gyration period is independent of the energy of the particle and is a function of the mass/charge ratio and the magnetic field only. In general

$$T_c = 2\pi \frac{m}{qB} \text{ which gives for}$$

$$\text{Protons: } T_c (\text{sec}) = \frac{6.51 \times 10^{-8}}{B(\text{Tesla})} \quad \text{or} \quad T_c (\text{sec}) = \frac{6.51 \times 10}{B(\text{nT})}$$

$$\text{Electrons: } T_c (\text{sec}) = \frac{3.5 \times 10^{-11}}{B(\text{Tesla})} \quad \text{or} \quad T_c (\text{sec}) = \frac{3.55 \times 10^{-2}}{B(\text{nT})}$$

Taking $B = 10^4$ nT as a typical value in the radiation belts we see that proton gyroperiods are typically in the millisecond range and electrons in the microseconds.

4 McIlwain L and Invariant Latitude

As we move into the magnetosphere, we need to define a particularly useful coordinate system. This system makes use of the fact that particles move along magnetic field lines, as developed above. Carl McIlwain developed the "L" parameter to provide a handy label for the magnetic field lines, and the McIlwain L-value is used throughout space physics. The L-value for a given magnetic field line is determined by (equal to) the distance from the center of the earth to the point at which it crosses the magnetic equator. The complementary parameter which is then needed to locate a region of interest is the magnetic latitude, λ_m . The ordered pair, (L, λ_m) , along with local time or longitude, then uniquely determines a location in space. The units of "L" are earth radii, magnetic latitude is typically given in degrees, and local time is generally preferred for the third coordinate. Thus the line which crosses the equator at $r = 3$ earth radii is labeled $L = 3$.

It can be shown that the radial distance, r , and latitude, or more properly, magnetic latitude (λ_m) are related according to the form:

$$\frac{r}{R_E} = L \cos^2 \lambda_m \quad (\text{Eqn. 5.20})$$

Again, we distinguish between the latitude, λ , and the magnetic latitude, λ_m .

The above expression is useful for all values of $\lambda_m < \Lambda_c$. Λ_c is the latitude at which a given line reenters the earth. This maximum latitude is given by

$$\cos \Lambda_c = \frac{1}{\sqrt{L}} \quad (\text{Eqn. 5.21})$$

Λ_c , or simply Λ is termed the invariant latitude, since, like L, it does not vary along the field line. In general, geophysicists specializing in the ionosphere use Λ , magnetospheric specialists use L.

The magnitude of the magnetic field is now given by:

$$B = \frac{B_{os}}{L^3} \frac{\sqrt{4 - 3 \cos^2 \lambda_m}}{\cos^6 \lambda_m} \quad (\text{Eqn. 5.22})$$

where:

B_{os} = Equatorial Field at the surface (equator)

Example:

How does the magnitude of the magnetic field vary with latitude (λ_m) along a magnetic field line, that is along a line of constant "L", or Λ . Beginning with equation 5.21, we use the values:

$L = 5.6$, $\Lambda = 65^\circ$, and assume $B_{os} = 30,000$ nT

we obtain the values:

$$B(L=5.6, \lambda_m=0^\circ) = 1.00 (B_{os}/176) = 170 \text{ nT}$$

$$B(\lambda_m=30^\circ) = 3.14 (B_{os}/176) = 534 \text{ nT}$$

$$B(\lambda_m=45^\circ) = 12.65 (B_{os}/176) = 2,150 \text{ nT}$$

$$B(\lambda_m=60^\circ) = 115.4 (B_{os}/176) = 19,618 \text{ nT}$$

Increase B as function of λ_m for the same L value, not the same r.

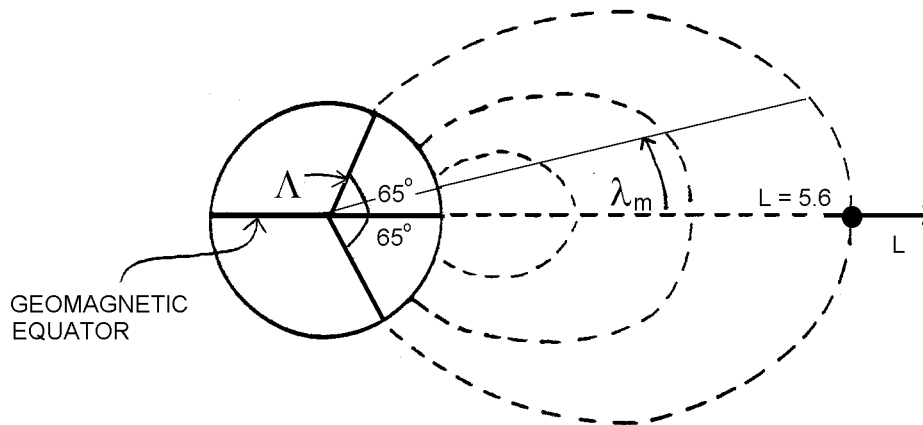


Figure 5.10 Magnetic field lines reach the surface of the earth at a latitude termed the Invariant Latitude.

C Magnetosphere structure

Returning now to the magnetosphere, we find the main features of the magnetosphere are (see Fig 5.1)

- (1) Bow Shock: Where the initial deceleration and deflection of solar wind particles occurs
- (2) Magnetopause: The outer boundary of the magnetosphere which divides magnetosphere and solar wind plasma.
- (3) Neutral Sheet: Extends behind the earth in the antisolar direction. Divides the geomagnetic tail into lobes with B lines pointing toward and away from the earth.
- (4) Plasma Sheet: A region of moderate particle density surrounding the neutral sheet. The plasma sheet is magnetically connected to the auroral ovals.
- (5) Trapped Radiation Belts: A region in which high energy electrons & protons gyrate around closed magnetic field lines between mirror points.
- (6) Plasmasphere: A region of relatively cold plasma (energies < 1 eV) which corotates with the earth due to frictional coupling.
- (7) Cusps: Regions which divide the magnetic field lines which close on the earth surface (sunlit side) from those which are stretched out into the tail.

Next let us consider some of these in a little more detail, starting with the bow shock which is the first thing the solar wind runs into as it approaches the earth.

D Bow Shock:

The observed shock wave in the solar wind is the result of a collective plasma interaction, usually referred to as a collisionless shock. It has some similarities to the aerodynamic shock formed around blunt objects but the mechanism by which the individual particles interact is not the same. In the aerodynamic case it is the actual collision of gas particles and the shock thickness is determined by the mean free path between collisions.

In the solar wind the mean free path is on the order of 10^8 km which is much larger than the radius of the earth. In fact the bow shock is not a collisional aerodynamic shock but is due to electromagnetic interaction between the magnetized solar wind (frozen-in field lines) and the magnetosphere. The aerodynamic formulas work reasonably well because the frozen-in field lines limit the particle motion of the solar wind plasma perpendicular to the field. This constraint makes the otherwise collisionless solar wind act like a collision dominated fluid (at least as far as motion perpendicular to the field is concerned).

We can estimate the thickness of the shock to be on the order of the gyration radius in the solar wind

$$r_c = \frac{mv}{qB} = \frac{(1.67 \times 10^{-27})(5 \times 10^4)}{(1.6 \times 10^{-19})(10^{-8})} \cong 50 \text{ km}$$

Where we have assumed protons of mass 1.67×10^{-27} kg with a transverse speed of 5×10^4 m/s which corresponds to a temperature in the solar wind of 10^5 °K. We have also assumed a local magnetic field of about 10 nT which agrees with measured values.

We thus find that the bow shock has a thickness on the order of 50-100 km.

It can also be shown that the ratios of particle velocity and particle density are approximately those predicted by theory, namely

$$V_{\text{postshock}} = \frac{1}{4} V_{\text{pres shock}} \quad \text{and} \quad \rho_{\text{pres shock}} = \frac{1}{4} \rho_{\text{postshock}} \quad (\text{Eqn. 5.23})$$

E Magnetopause

We may look upon the magnetopause as the boundary between the region dominated by the geomagnetic field on one side and the region dominated by the solar wind plasma pressure on the other side. We can make an order of magnitude calculation to show where this boundary will be located under normal solar wind conditions: We have seen earlier that the quantity $(B^2/2\mu_0)$ can be considered as a magnetic pressure term.

We can therefore write the condition for the location of the magnetopause by equating the total pressure on one side to the total pressure on the other side

$$P_1 + \frac{B_1^2}{2\mu_0} = P_2 + \frac{B_2^2}{2\mu_0} \quad (\text{Eqn. 5.24})$$

Next we assume that $\frac{B_1^2}{2\mu_0} = P_2 = 0$ which says that there is only particles on the left side and only field pressure on the right side.

The pressure is not nkT in this case, because the energy of the solar wind is a directed energy, and the resulting pressure is:

$$P_1 = nmv^2 \quad (\text{Eqn. 5.25})$$

where:

- n = particle density
- m = particle mass
- v = incident particle speed

Thus

$$nmv^2 = \frac{B_2^2}{2\mu_0} \quad (\text{Eqn. 5.26})$$

We can also show that the compressed value of B, that is B_2 , is equal to twice the uncompressed dipole B value.

$$B_2 = 2 \frac{B_{os}}{\left(\frac{r}{R_{\oplus}}\right)^3} \quad (\text{Eqn. 5.27})$$

where $\left(\frac{r}{R_{\oplus}}\right)_s$ = standoff distance in earth radii

$$4 \frac{B_{os}^2}{\left(\frac{r}{R_{\oplus}}\right)^6} = 2\mu_0 nmv^2$$

where B_0 = Equatorial surface field strength. Solving for r gives:

$$\left(\frac{r}{R_{\oplus}}\right)_s^6 = \left(\frac{4 B_0^2}{2\mu_0 nmv^2}\right)$$

can be written as

$$\left(\frac{r}{R_{\oplus}}\right)_s = \left(\frac{4 B_{os}^2}{2\mu_0 nmv^2}\right)^{1/6} = \text{stand off distance} \quad (\text{Eqn. 5.28})$$

Plug in some reasonable values:

$$B_{os} = 3.0 \times 10^{-5} \text{ T}, \quad n = 10^7 \text{ m}^{-3}, \quad m = 1.67 \times 10^{-27} \text{ kg}, \quad v = 4 \times 10^5 \text{ m/s}$$

$$\left(\frac{r}{R_{\oplus}}\right)_s = \left(\frac{4 \times (3 \times 10^{-5})^2}{(10^7)(1.67 \times 10^{-27})(4 \times 10^5)^2(8\pi \times 10^{-7})}\right)^{1/6} = 9.01$$

Thus under these typical conditions the standoff distance is about 9 earth radii, a value which is in tolerably good agreement with observations.

An interesting feature of this result is the fact that as a result of the (1/6) exponent even rather large changes in the solar wind will not affect the standoff distance a great deal. A change of a factor of 10 in the product nmv^2 will only change $(r/R_{\oplus})_s$ by about a factor of 1.5 and that is a pretty extreme case. Thus the magnetopause is a surprisingly static boundary.

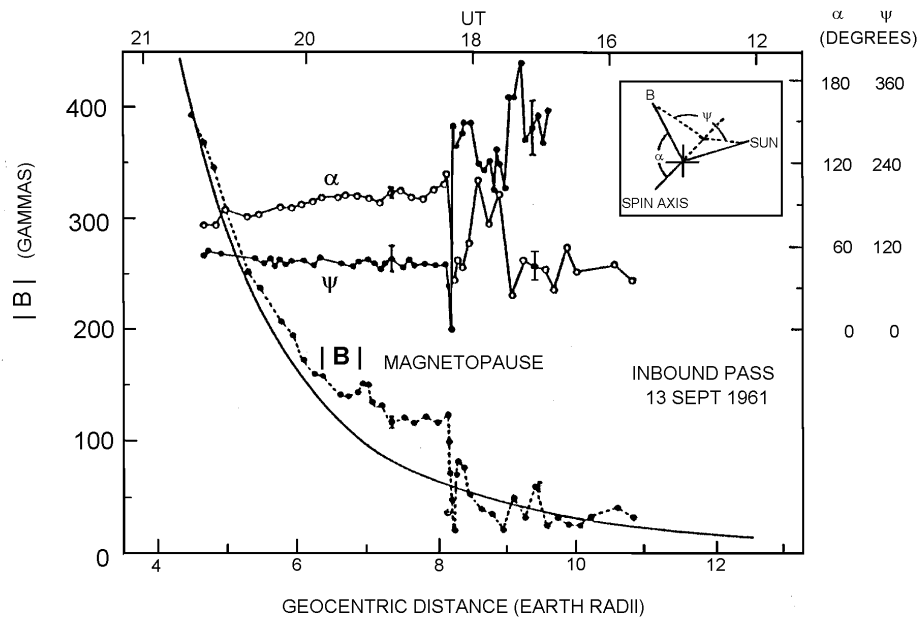


Figure 5.11 A magnetometer record of the geomagnetic field obtained near the noon meridian during an inbound pass of Explorer 12 on September 13, 1961. The magnitude of the total magnetic intensity $|B|$ and the angles α and ψ , defined in the inset figure, are plotted against geocentric distance from the earth ($1 \text{ gamma} = 10^{-5} \text{ Gauss} = 1 \text{ nT}$). The smooth curve shows the variation of $|B|$ with distance for the *Finch and Leaton* [1957] geomagnetic reference field.

From: Cahill and Amazeen, The boundary of the geomagnetic field, in *Journal of Geophysical Research*, 68, 1835-1843, 1963, or in *Planetary and Space Science*, 15, 997-1033, 1967.

Found in review article: Structure of the Magnetopause, D. M. Willis, *Reviews of Geophysics and Space Physics*, 9, 953, 1971.

F Geosynchronous Orbit - The Plasmasphere and the Plasma Sheet

Within the magnetosphere are two regions of primary interest (as far as satellites in geosynchronous orbit are concerned). They are the plasma sheet and plasmasphere. Satellites in geosynchronous orbit will generally be in one of these two regions, depending upon the local time and the level of magnetic activity. The relative locations of these regions are partially indicated in figures 4.1 and 5.1, and are shown in cross-section views in figure 5.12. Note that the Navstar (GPS) satellites, in their $4 R_E$, circular, inclined orbits, will pass through these same plasma regions.

The plasmasphere is generally thought of as the upward extension of the ionosphere, a region of relatively cold, dense plasma. The outer boundary of the plasmasphere is the plasmapause, typically found from 4-6 earth radii out, as observed at the equator. Outside this boundary region we find the plasma sheet, a region of relatively hot, tenuous plasma. Geosynchronous satellites are typically in the plasmasphere from ~ mid-afternoon (1500 Local time) to early evening (~2000 LT), and in the plasma sheet from early evening to just past dawn (~0600 LT). The region from dawn to mid-afternoon is an ambiguous region of mixed plasmas.

The region between the plasmasphere proper and the plasma sheet, as observed on the night side of the earth, is nominally the plasmapause region, which of necessity has a finite thickness. The gap between these two regions, as given by Vasyliunas, is partly an artifact of the instrumentation available at that time, but still accurately reflects the somewhat ambiguous nature of the mixed regions of hot and cold plasma just outside the plasmasphere.

The plasmasphere has as its low altitude boundary the ionosphere, a boundary which is nominally at an altitude of 1000 km, though the distinction is not a clear one. The plasma sheet begins at 5-6 earth radii at the equator, but note that the hot plasma follows the magnetic field lines up to a high latitude intercept with the ionosphere, at magnetic latitudes from 60-70°.

There is a high latitude limit, corresponding to magnetic field lines extending out 10's to 100's of earth radii, depending upon magnetic activity.

As noted above, the plasmasphere is a region of cold, 'dense' plasma near the earth, extending out to geosynchronous orbit at times. The density of the plasma inside the plasmasphere ranges from

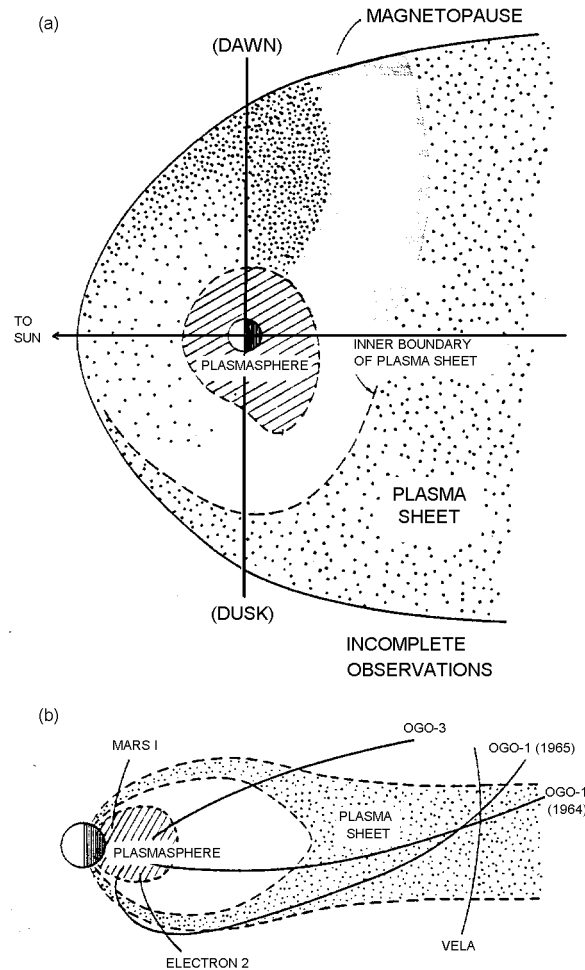


Figure 12. Diagram of the plasmasphere and plasmasheet. Illustration shows the average distribution of low-energy electrons (a) in the equatorial plane, and (b) in the meridian plane (noon-midnight slice) of the magnetosphere. Orbits of significant early satellites are indicated. Here, low energy electrons means a few hundred eV. Figure from Vasyliunas, V. M. (1972).

Vasyliunas, V. M., *Magnetospheric Plasma*, in *Solar Terrestrial Physics 1970*, E. R. Dyer and J. G. Roederer, eds., D. Reidel Publishing Co., Dordrecht, Holland, 1972. A survey of low-energy plasma measurements in space.

a peak value of $\sim 10^4/\text{cm}^3$ down to $\sim 1 \text{ cm}^{-3}$. Data from OGO-5 show a variety of plasma (ion) density profiles obtained in 1968 (Figure 5.13). It can be seen that there is a gradual slope at low "L", followed by a sharp drop in density. The gradual slope is a reflection of the increasing volume of the magnetic field 'flux' tubes, and it is often found that the total plasma density is inversely proportional to the fourth power of L.

$$n (\text{cm}^{-3}) = 100 \left(\frac{L}{4.5} \right)^{-4} \quad (\text{Eqn. 5.29})$$

The plasmasphere ions and electrons are 'cold' with characteristic temperatures of 0.5 - 1.0 eV. There are equal numbers of ions and electrons (the quasi-neutrality condition), and the ions are primarily H^+ and He^+ , with a smaller percentage (usually) of O^+ , O^{++} , N^+ , He^{++} , and N^{++} .

The plasmasphere is said to 'co-rotate' with the earth. The cold plasma is well tied to the earth's magnetic field. Corresponding to this motion is an electric field, just sufficient to produce the observed $\mathbf{E} \times \mathbf{B}$ 'drift'. At any rate, this effect drops off outside the plasmasphere, more or less near geosynchronous orbit.

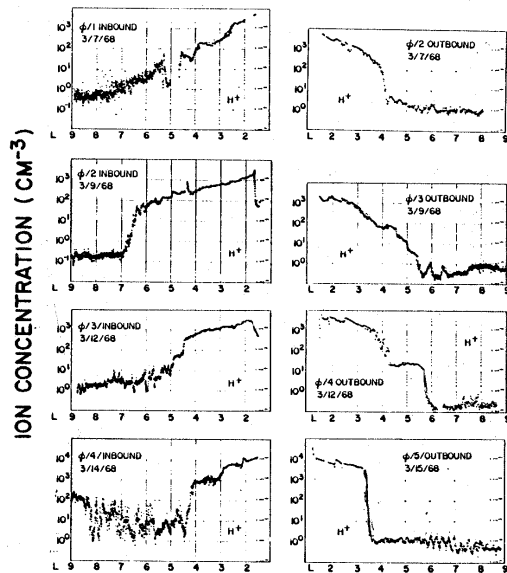


Figure 5.13 H^+ density distribution ($n_i(L)$) profiles from ion mass spectrometer data obtained on board the OGO-5 satellite. Here, L measures the equatorial distance in earth radii $R_E = 6375 \text{ km}$ (from Harris et al, 1970)

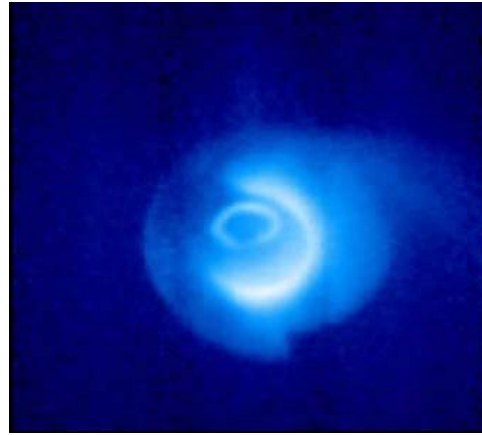


Figure 5.14 Image satellite observations of He^+ (HE II) at 304 \AA . (Not unlike viewing the sun from SOHO – just not nearly as bright!) IMAGE was launched March 25, 2000. Image from May 2000? Satellite near apogee – 6-7 R_E .

The plasma sheet is the region of hot plasma ($T \gg 1$ keV) with a typical density of $\sim 1/\text{cm}^3$. During magnetic 'substorms' the temperature of the plasma sheet electrons can reach 10-20 keV, and substantial fluxes of electrons are seen at energies up to 50-100 keV. The ions which make up the plasma sheet generally show about a 50-50 mixture of singly charged hydrogen (H^+) and oxygen (O^+). These ions are thought to come from the upper atmosphere of the earth (because we see O^+ , a terrestrial ion, as opposed to He^{++} or O^{6+} , solar wind ions).

The plasma sheet is typically thought of as being a feature of the night-side of the magnetosphere, particularly the midnight-to-dawn sector, though the hot plasma can drift around the earth into the dawn-noon sector. It extends from $L \sim 5$ or 6 to beyond $L=20$. the location of the inner boundary depends on magnetic activity. During a period of high activity (high K_p , and many substorms), the plasma sheet may extend in to $L = 3$ or 4, and will be seen at geosynchronous orbit from near dusk, through midnight, and on past dawn.

Modern data from a particle detector on an operational geosynchronous satellite are shown in figure 5-15. The satellite is in the plasmasphere from 0-4 UT, then moves into the plasma sheet for the rest of the day. Local midnight occurs at 1100 UT. This is roughly where the hottest (most energetic) electrons are seen. Note that this is a fairly quiet period, but the electrons observed at local midnight extend to many keV.

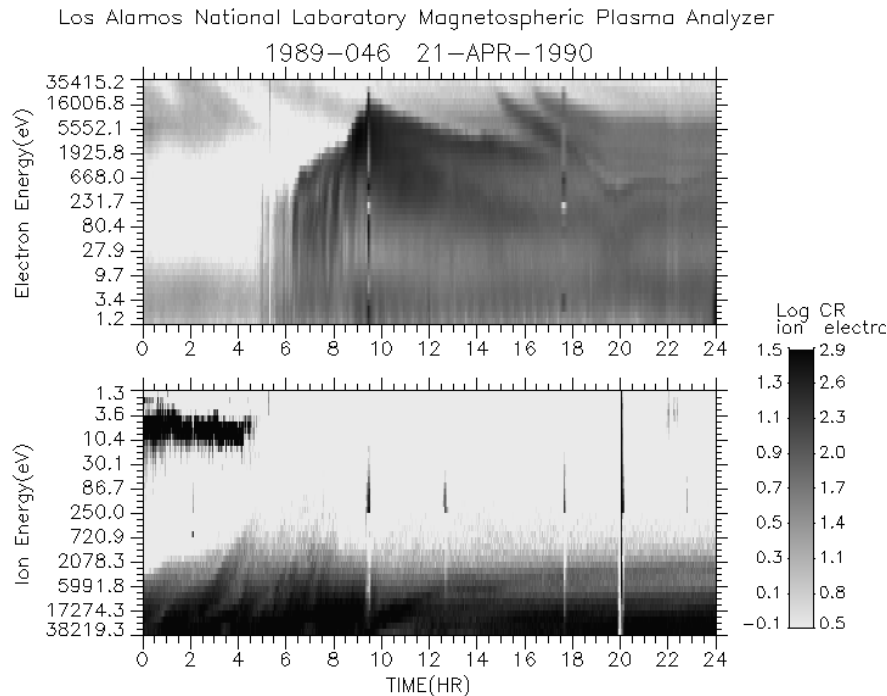


Figure 5-15 Ion and electron observations from the LANL charged particle analyzers on the geosynchronous satellite, 1989-046. Data are plotted such that low energy electrons are at the lower boundary of the top panel, while the ion data are inverted, with lowest energies at the top. This facilitates the comparison of the lower energy electron and ion data. High fluxes of low energy ions (3-10 eV) are seen from 0-4 UT, indicating that the satellite is in the plasmasphere. Relatively high fluxes of more energetic electrons begin to appear at 0600 UT, roughly dusk local time.

G Magnetospheric Convection

Why do the data at geosynchronous orbit appear as they do? A reasonable explanation for the patterns observed was obtained in the first decade of space plasma physics, and the theory has remained reasonably consistent since that time. The answer lies in magnetospheric convection and drifts. The concept was nominally due to Neil Brice, at NASA/GSFC, and has been termed Brice convection. Atsushiro Nishida developed the concept in an early classic text on the topic.

A) Because of the flow of the solar wind past the magnetosphere, and the processes being driven at the magnetopause, there is imposed on the magnetosphere an electric field in the dawn-dusk direction with a magnitude of ~ 1 mV/m. This electric field is such as to force plasma sunward from the tail, a region where we believe hot plasmas are created. Note that integrated over the magnetosphere, there is a 50-80 kV potential drop associated with this electric field.

B) Because the inner magnetospheric region called the plasmasphere is corotating with the earth, there must be a radial electric field (inward) associated with the plasma rotation. For zero energy particles, the balance between these two competing fields determines a set of boundaries, most easily determined by calculating the equipotential surfaces.

Following Parks (1991), we obtain the corotation field in the equatorial plane:

In spherical coordinates:

$$\vec{U} = \omega r \hat{\phi} \quad (\text{Eqn. 5.30})$$

$$\mathbf{B} = B_0 \left(\frac{R_{\oplus}}{r} \right)^3 (-\hat{\theta}) \quad (\text{Eqn. 5.31})$$

where U is the streaming, or drift velocity. The frozen-in-field condition gives:

$$\begin{aligned} \mathbf{E}_{\text{corotation}} &= -\mathbf{U} \times \mathbf{B} = \omega r \hat{\phi} \times B_0 \left(\frac{R_{\oplus}}{r} \right)^3 (\hat{\theta}) \\ &= \frac{\omega B_0 R_{\oplus}^3}{r^2} (-\hat{r}) \end{aligned} \quad (\text{Eqn. 5.32})$$

The corresponding potential is easily shown to be:

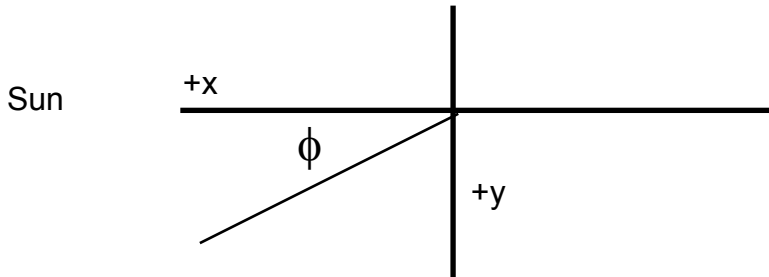
$$\Psi_{\text{corotation}} = -\frac{\omega B_0 R_{\oplus}^3}{r} = -\frac{\omega B_0 R_{\oplus}^2}{L} \quad (\text{Eqn. 5.33})$$

The cross-tail electric field can be expressed as:

$$\mathbf{E}_{\text{sw}} = E_{\text{sw}} \hat{y}, \quad \text{where } \hat{x} \text{ is sunward, } \hat{z} \text{ is north,} \quad (\text{Eqn. 5.34})$$

and hence \hat{y} is from dawn to dusk

$$\Psi_{\text{sw}} = -E_{\text{sw}} y = -E_{\text{sw}} r \sin \phi = -E_{\text{sw}} L R_{\oplus} \sin \phi \quad (\text{Eqn. 5.35})$$



Adding the two potentials together, we obtain

$$\Psi_{\text{total}} = -E_{\text{sw}} L R_{\oplus} \sin\phi - \frac{\omega B_0 R_{\oplus}^2}{L}$$

The resulting equipotential contours are illustrated in Figure 5.16. It can be seen that there are two distinct regions. In the region near earth, the equipotential lines are closed, and ~circular. These are streamlines for the drift of ~zero energy particles (no significant ∇B drift). Outside the curiously pointed boundary crossing the duskward axis at $L=7.5$, the streamlines are 'open', meaning the plasma flowing along these lines from the tail will pass by the earth, terminating at the magnetopause.

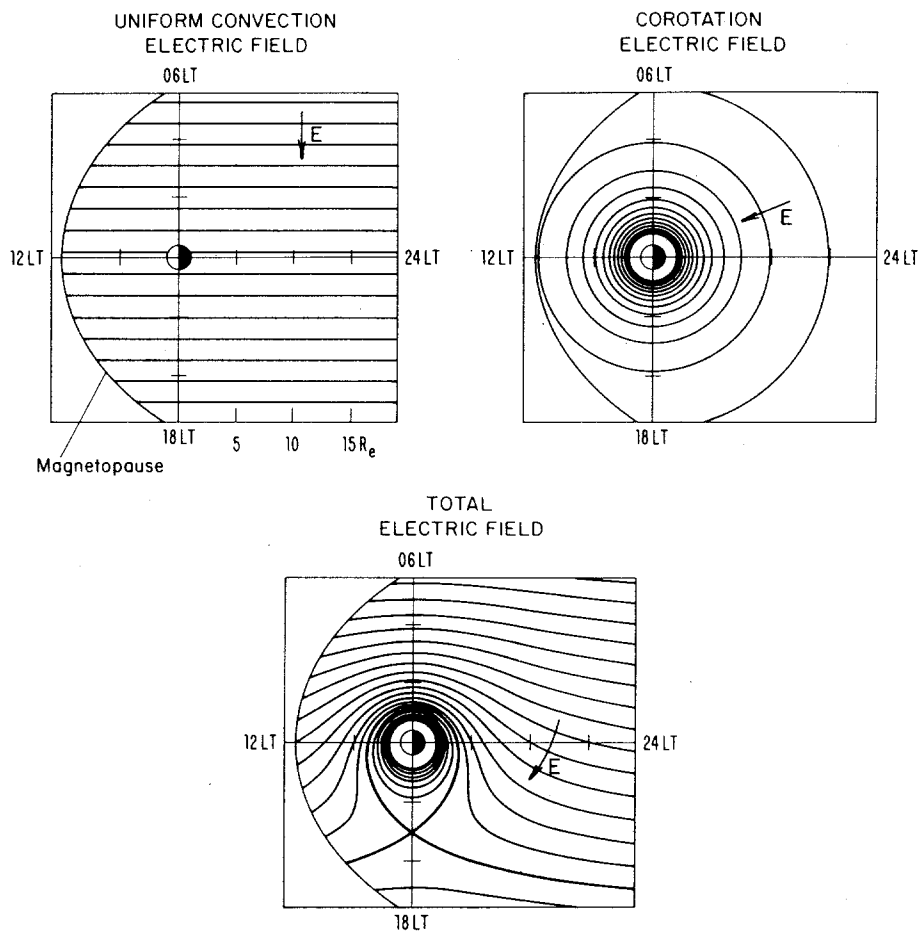


Figure 5.16 Lyons, L.R., and D. J. Williams, *Quantitative Aspects of Magnetospheric Physics*, D. Reidel Publishing Co. Boston, MA, 1984

Found in: George K. Parks, *Physics of Space Plasmas*, Addison-Wesley Publishing Co., Redwood City, CA, 1991, page 234.

A formula for the location of this boundary can be obtained by recognizing that it is the boundary between the region where the solar wind induced, cross-tail field dominates, and the region where corotation dominates. Dealing just with the magnitudes at 1800 LT:

$$E_{\text{corotation}} = \frac{\omega B_o R_{\oplus}}{L^2} = E_{\text{SW}}, \text{ or } L_{\text{Plasmapause}} = \sqrt{\frac{\omega B_o R_{\oplus}}{E_{\text{SW}}}} \quad (\text{Eqn. 5.36})$$

$$L_{\text{plasmapause}}(1800) = \sqrt{\frac{\omega B_o R_{\oplus}}{E_{\text{sw}}}} = \sqrt{\frac{2\pi}{86400} \frac{3 \times 10^{-5} \cdot 6.37 \times 10^6}{0.35 \times 10^{-3}}} = 6.3 \quad (\text{Eqn. 5.37})$$

If particles with significant kinetic energy are considered, then the grad-B drift must be included, at least near the earth. The grad-B drift is azimuthal, and for electrons, is in the same direction as co-rotation. This term has the effect of expanding the corotation equipotentials - and the separatrix moves outward.

For ions, the grad-B drift opposes the corotation field, and for sufficiently high energies (above ~ 5 keV at geosynchronous orbit), they will counter-rotate. In close to the earth, it is the grad-B drift of the ions which produces the ring-current.

H The Radiation Belts

1 The Van Allen Radiation Belts

The (Van Allen) radiation belts, are of primary concern for the satellite designer because of the effects of penetrating radiation on electronics. The Van Allen belts permeate the plasmasphere and inner plasma sheet, but are generally thought of as co-existing with the plasmasphere. It is the particles of the radiation belts that primarily form the ring current.

Figure 5.17 shows the ion flux vs. L , with ion energy as a parameter. For purposes of study, focus on two energies which illustrate the main points here. Ions of 1 MeV energy have a peak flux at $L \sim 3$. This would be the nominal location of the canonical 'ring current'. The L value at which the peak flux occurs decreases as energy increases. The 400 MeV ions, for example, have a peak flux at $L \sim 1.5$. (These plots are based on the standard NASA model held at the National Space Science Data Center, and is commonly attributed to Jim Vette.)

(<http://nssdc.gsfc.nasa.gov/space/model/magnetos/aeap.html>)

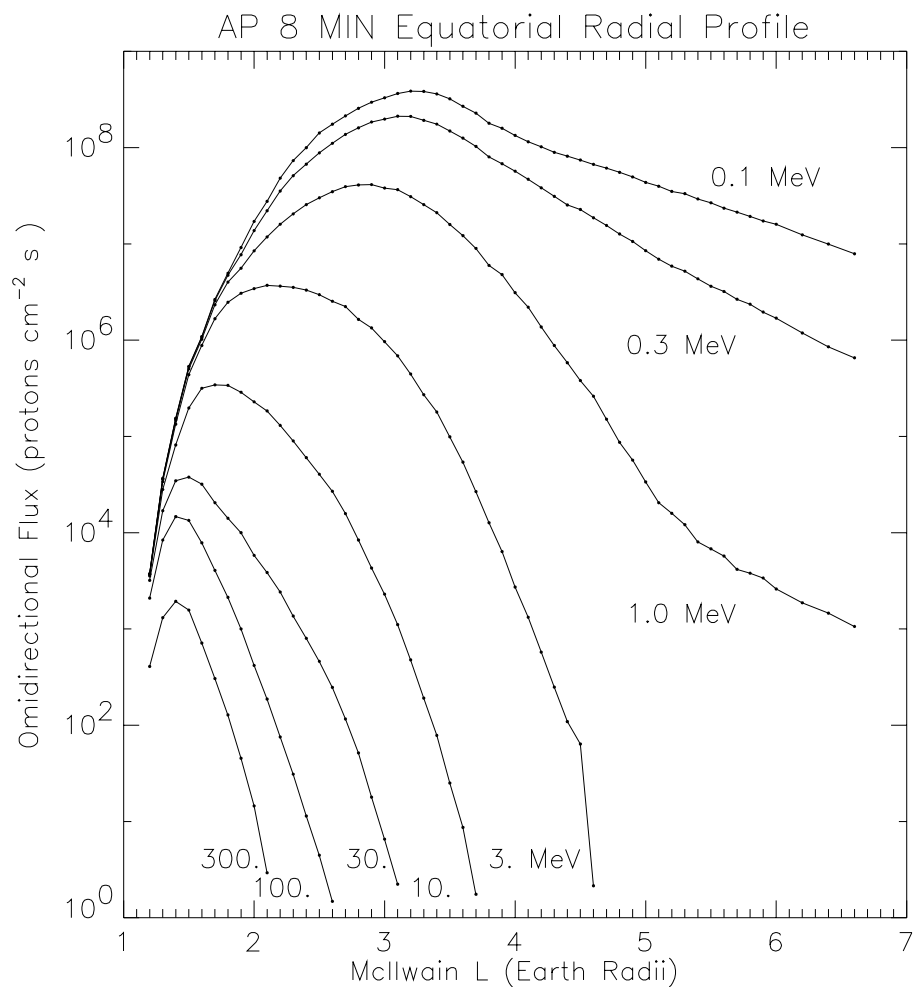


Figure 5.17 Radiation belt ion flux vs. L. The plotted values are from the AP8MIN model for omnidirectional flux of protons (ions) in the equatorial plane with energies above threshold values between 0.1 and 400.0 MeV [Sawyer and Vette, 1976][<http://nssdc.gsfc.nasa.gov/space/models/trap.html>].

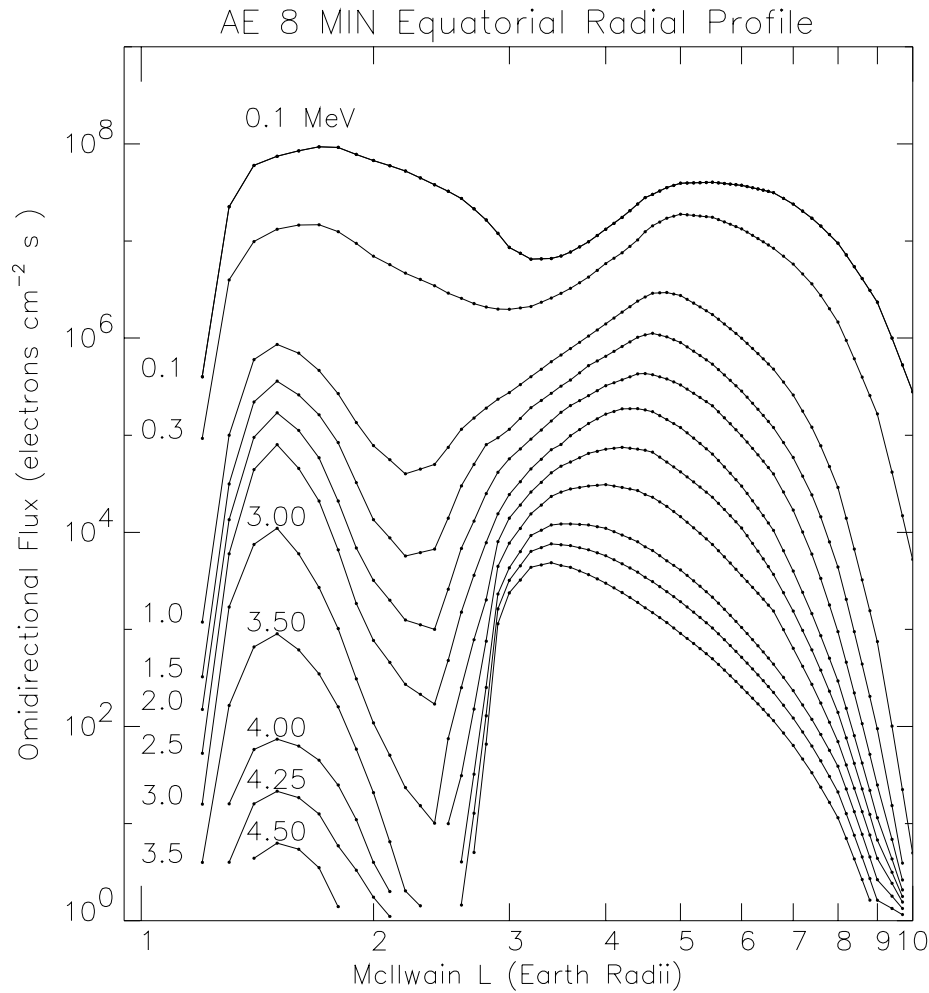


Figure 5.18 Radiation Belt Electrons - Solar Minimum. The curves are labeled with the electron energy in MeV, ranging from 0.1 to 4.5 MeV. Note the logarithmic horizontal axis. Graphed data obtained from the National Space Science Data Center, at NASA/GSFC. <http://nssdc.gsfc.nasa.gov/space/models/trap.html>

Figure 5.18 shows the electron flux vs. L , with electron energy as a parameter. Again we focus on two energies which illustrate the main points here. Electrons of 1 MeV energy have a peak flux at $L \sim 1.5$, coinciding with the so-called, 'inner belt', and the location of the highest energy ions. There is a second peak at $L \sim 5$, coinciding with the more traditional ring current, and (relatively) lower energy ions.

Given the above characteristics for the average condition of the radiation belts, we again come to the question of why they take the form shown above, and what are the significant variations about the model values shown so far. It is now also appropriate to see how the radiation belts are related to the phenomenon observed in ground magnetometer data, in particular, the ring current.

2 The Ring Current

Since the ring current was originally described to you as an inference from the magnetic storm data (e.g. Dst and discussion), we return briefly to that point of view. First of all, does the observed distribution of energetic particles explain the variation in magnetic field strength ascribed to a 'ring current'? The answer is yes.

The magnetic field perturbation produced by the particles extends into space, and suitably clever manipulation of the observations allows one to show that the observed particles are capable of producing the observed perturbation in magnetic field. Note that the current is primarily produced by grad-B drift of protons.

What are these ring current ions, and where do they come from? Many of the plots you see, are generally labeled 'protons', but are really for all ions. This is because most of the early instruments, and many of the current ones, such as NOAA-10 detectors, do not discriminate between ions. It is desirable to do so, however, so that one can sort out where these ions come from, and then understand the formation and decay of the ring current.

If we look again at a typical magnetic storm signature (Dst), we find two main puzzles. Where do the particles come from, and where do they go. First, let's concentrate on the decay of the ring current, also termed the storm recovery phase. We start there, because it's better understood.

Two concepts come up, at this point, which have been presented above:

- a) there are low energy charged particles for the ring current to interact with, e. g. the plasmasphere, and
- b) there is a neutral gas background to interact with, as revealed by the geocorona (addressed next).

The low energy charged particles of the plasmasphere (described above) can interact with the ring current via the medium of electromagnetic waves (photons). A particularly important subset of such waves are the waves (radio signals) at a few kilo-Hertz frequency, known as whistlers. These waves are generated, initially, by lightning in the atmosphere. The waves travel into space, where they can scatter off the charged particles (photon collides with particle, transfers momentum, changes pitch angle). If the particle is knocked into the loss cone, it will fall into the atmosphere, and be lost. This works particularly well for the high energy electrons. It is less important for ions. Note that because there are lots of low-energy electrons and ions in the plasmasphere; it doesn't affect charge neutrality if a few of the MeV particles are knocked out.

Ions are traditionally thought of as being lost via charge-exchange with the neutral hydrogen background. This background is revealed in the observations of neutral hydrogen made in the UV (Lyman-alpha) of the near-earth region.

3 The geocorona

The uppermost region of the neutral atmosphere is a region of neutral hydrogen gas extending well into near-earth space. This gas is visible in reflected UV sunlight in the Lyman- α emission (scattering). This light is termed the geocorona. Figure 5.19a shows the geocorona, as observed from the Moon by Apollo astronauts (taken with an NRL camera). This light is a visible indication of the neutral hydrogen background which is effectively the top of the earth's atmosphere. The glow can be modeled, and from that the neutral hydrogen density profile can be determined. Figure 5.19b shows a model developed by Rairden, using data from the DE-1 satellite. The neutral hydrogen density at geosynchronous orbit is about $10 / \text{cm}^3$ - about the same as the plasma density, depending upon local time.



Figure 5.19a Lyman-alpha glow of the geocorona photographed from the moon. The earth-centered cloud reaches 50,000 miles, and its intensity exceeds the total of all visible airglow.

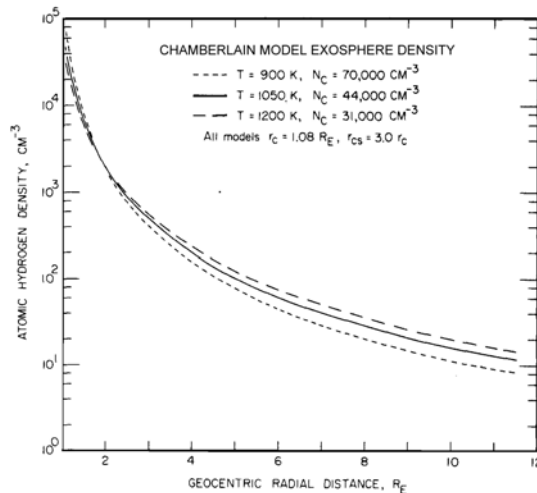


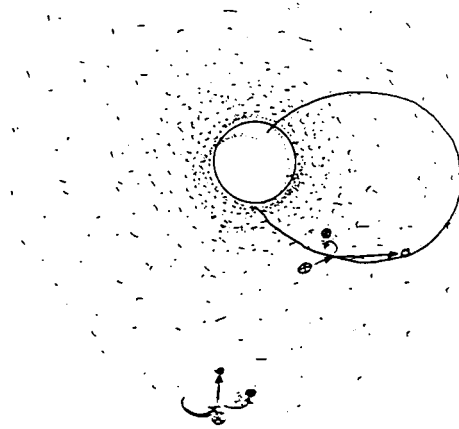
Figure 5.19 b. Exospheric hydrogen density versus radial distance for the Chamberlain model. The model at temperature $T = 1050 \text{ K}$ provides the best fit to the DE 1 geocoronal observations.

From: Geocoronal imaging with Dynamics Explorer, R. L. Rairden, L. A. Frank, and J. D. Craven, *Journal of Geophysical Research*, vol. 91, page 13613, 1986.

4 Ring Current Decay and Composition

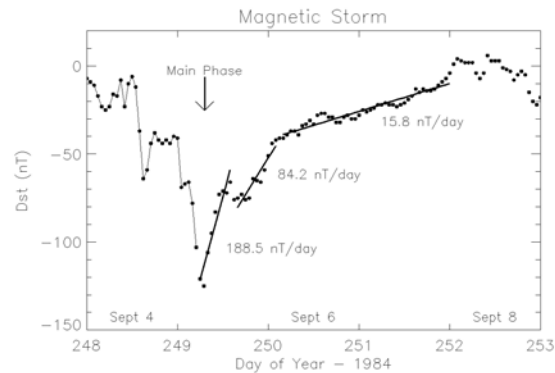
This neutral gas background provides a scattering mechanism for the energetic ions of the radiation belts. This process is illustrated in Figure 5.20. An energetic ion will collide with a low energy neutral particle, and take the electron away. There is now an energetic neutral atom, which heads off towards Pluto, and one more (boring) low energy ($\sim 0.1 \text{ eV}$) proton. This is a well understood physical process, and the time scale for decay via charge exchange was one of the first quantitative calculations to be done.

Figure 5.20 During charge exchange a **hot** (or energetic) ion collides with a **cold** neutral hydrogen atom, the same neutral hydrogen "visible" in the UV as the geocorona. The **hot** ion picks up an electron from the **cold** H atom and becomes a **hot** neutral particle, which is no longer trapped in the earth's magnetic field. The **cold** H, which lost its electron, is now trapped instead. Thus the process causes a loss of the energetic ions from the ring current. [This figure provided by Joe Fennell.]



These charge exchange calculations, which assumed that the ring current was composed only of protons, were inconsistent with some observations, however, of the type shown in figure 5.20. The time scales for decay of ring current ions indicated that there must be more than one kind of ion out there, and that the initial rapid recovery of the ring current is inconsistent with H^+ being the (initial) primary component of the ring current. We now understand that the ring current, as formed during magnetic storms, is composed of O^+ , He^+ , and H^+ . The heavier ions decay more quickly, as indicated by figure 5.21, indicating that there was indeed substantial amounts of oxygen (and helium) in the ring current. This helps address the question of where the ring current particles originate.

Figure 5.21 The recovery phase of the magnetic storm exhibits several characteristic time scales, corresponding to the different elements (ions) which make up the ring current. These time scales are revealed by the 3 characteristic slopes found during the storm illustrated here. In sequence, they reveal the decay of the O^+ , He^+ , and finally H^+ . The 3 time scales are due to the different cross-sections for the collision process of charge-exchange.



Where does the ring current come from ? What is it made of ? Where does it get all that energy ?

Initially, people thought the ring current ions must come from the sun, since the earth could not (apparently) accelerate ions and electrons to the observed energies. Observations of the mass composition of the ring current show this to be incorrect. Figure 5-21 shows the relative abundance of several oxygen ions in the ring current during the storm illustrated in Figure 5-20. This is the first direct measurement of ring current composition. Oxygen was found to be the major ring current component of the initial phase of the storm, in singly charged form. Where does it come from ? The ratio of O^+ (and O^{++}) to O^{6+} (oxygen stripped of 6 electrons) show that ions which originate in the earth's atmosphere dominate those of solar origin. The ring current ions must come from the earth's atmosphere.

How are they energized ? We don't know. If you look in the magnetosphere for a power source, there is the cross-tail electric field, which is good for 100 keV or so. The major alternative is the dB/dt of a storm which should induce a good size voltage around the earth.

Figure 5.22 Charge-state histogram of 1 to 300 keV/e ring current oxygen during the Sept 4-7, 1984 geomagnetic storm. Measurements were made on the AMPTE/Charge Composition Explorer (AMPTE/CCE) satellite, with the CHEM experiment, built by the University of Maryland and the Max Planck Institut. Notice the *extremely* long accumulation interval.

From: First composition measurements of the bulk of the storm-time ring current (1 to 300 keV/e) with AMPTE-CCE, G. Gloeckler, B. Wilken, W. Stüdemann, F. M. Ipavich, D. Hovestadt, D. C. Hamilton, and G. Kremser, Geophysical Research Letters, vol. 12, page 325, 1985.

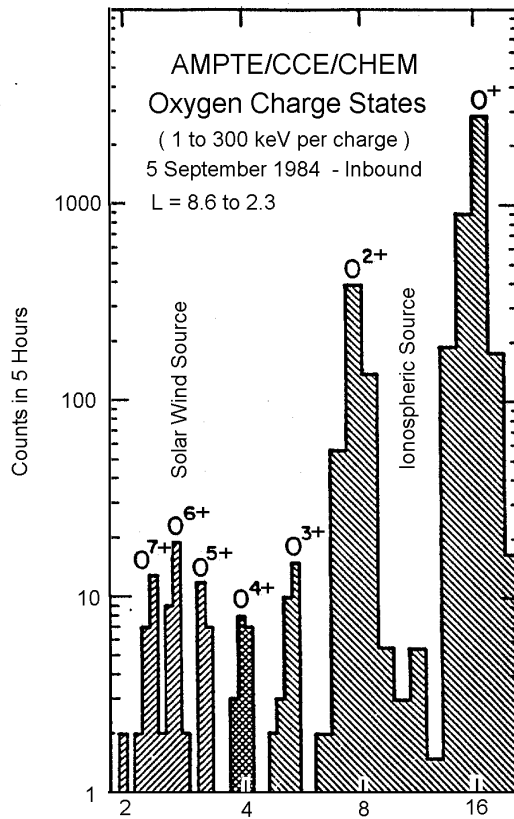
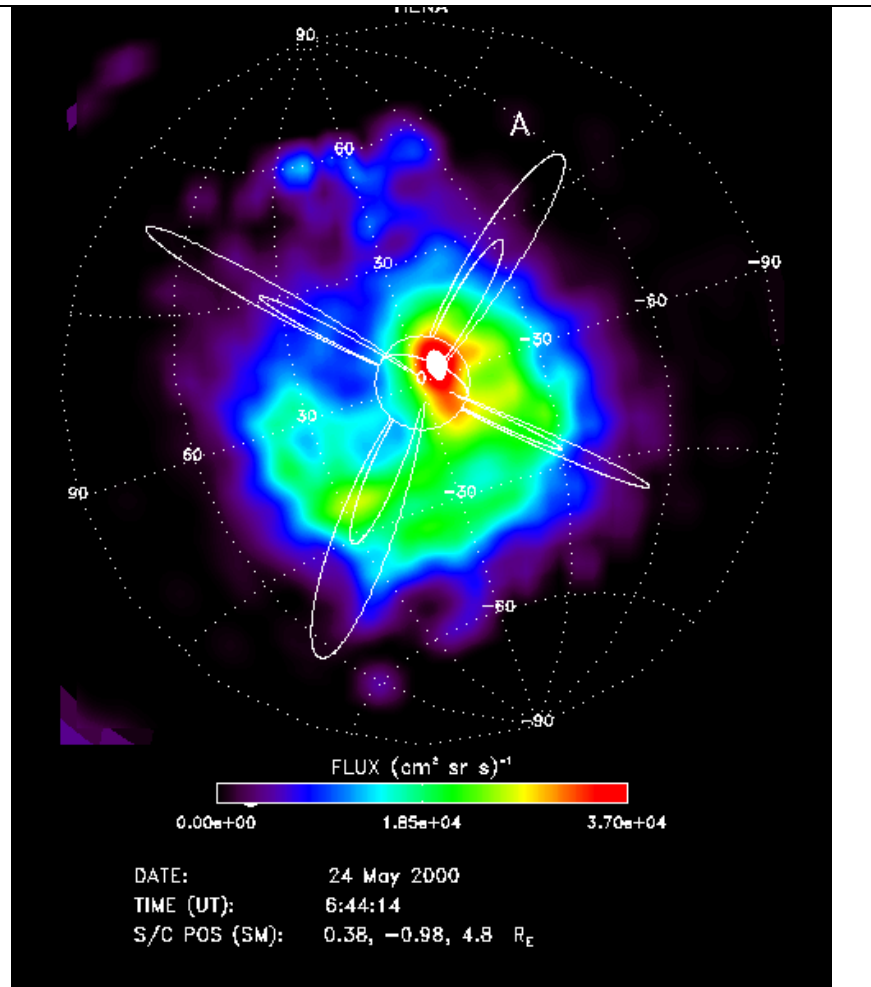


Figure 5.23

Recently, the radiation belts have been observed by the IMAGE satellite, using the charge-exchanged ions escaping from the magnetosphere. The satellite is viewing from the sun towards the Earth, viewing the cloud beyond the Earth on the night side.

This image is a convolution of the intensity of the radiation belts, and the density of the neutral hydrogen background.



I Problems

1. Calculate the magnetopause standoff distance for shock conditions - $v \sim 800$ km/s, $n = 25$ cm⁻³.
2. The electric field in the plasmasphere is just such as to cause the cold plasma to 'corotate' with the earth. What is the electric field necessary to produce this effect? (magnitude and direction) The answer is given in section 5.8.
Calculate and plot $E(r)$ from $L = 1$ to 6.
($B_0 = 3 \times 10^{-5}$ Tesla, $R_{\oplus} = 6.7 \times 10^6$ m)
3. Calculate the Invariant latitudes which correspond to $L = 4$, and $L = 6.6$.
4. Compare the L^{-4} electron density profile (Eqn. 5.29) with the Rairden article profile for neutral hydrogen density (Figure 18b). Do this by making a plot, and evaluating the two quantities at $\sim 1 R_{\oplus}$ intervals. Is there a point where there are more electrons (and presumably ions) than there are neutral H atoms?
5. For 1 MeV ions, at $L = 3$, calculate the grad - B drift velocity.
Assume a dipole magnetic field.
Assume the pitch angle (α) is 90° .

Note: $\nabla |\vec{B}| = \frac{\partial}{\partial r} \left\{ B_0 \left(\frac{R_{\oplus}}{r} \right)^3 \right\} = -3 \left(\frac{|\vec{B}|}{r} \right) \hat{r}$ in the magnetic equatorial plane,

hence $\frac{\nabla |\vec{B}|}{|\vec{B}|} = -3 \hat{r} / r$

6. Estimate the density for 1 MeV protons at $L = 3$. Use the model found in Figure 5-16, which shows the flux $\sim 3 \times 10^7$ (ions/cm² s). It might help to know that the flux, $f = nv$, where v is the average velocity. The average velocity can be found by solving the well known equation $K = \frac{1}{2} mv^2$, after carefully converting the energy from MeV to Joules, and the flux from ions/cm² s to ions/m² s. You pretty much have to assume the ions are protons ($m = 1.67 \times 10^{-27}$ kg)
7. Estimate the current density ($J = qnv$) due to 1 MeV protons, where the velocity is the grad-B drift velocity. By assuming this current density flows through a box 1 earth radius on a side, one can estimate a total current, $I = J \cdot \text{Area}$. Estimate I , and compare it to the current calculated in chapter 4 homework as being required to produce the observed Dst..

$$\begin{array}{c} \text{====>} \\ J = qn\vec{V} \end{array} \quad \begin{array}{c} | \\ | \\ | \\ | \\ | \end{array}$$

8. Use Faraday's law to calculate the electric field induced around the earth during a storm, ignoring the fact that much of this dB/dt is produced by the particles, themselves.

$$dB \sim 100 \text{ nT}, dT = 100 - 1000 \text{ s}, r = 3 R_E (20,000 \text{ km})$$

9. Evaluate the magnetic field strength for $L = 5$, $\lambda_m = 0^\circ, 45^\circ$, then calculate the critical (equatorial) pitch angle such that particles which will mirror at $\lambda_m = 45^\circ$.

10. Magnetospheric Convection: The plasmapause potential contour illustrated in Figure 5.15 can be obtained by returning to the convection discussion, and solving for the first closed contour.

a) Using, $L_{\text{plasmapause}}(\phi = 90^\circ) = \sqrt{\frac{\omega B_o R_\oplus^2}{E_{\text{sw}}}}$ (at 1800 LT),

solve for $\Psi_{\text{plasmapause}}$ (at $\phi=90^\circ$ or 1800 LT). (A formula which is a function of E_{sw})(Use Eqn. 5.33)

b) Solve the equation: $\Psi_t = -E_{\text{sw}} L R_\oplus \sin\phi - \frac{\omega B_o R_\oplus^2}{L}$ for L and show that the following relation can be obtained:

$$L = \frac{\Psi_t}{A} \pm \left[\left(\frac{\Psi_t}{A} \right)^2 - \frac{2\omega B_o R_\oplus^2}{A} \right]^{\frac{1}{2}}; \text{ where } A = 2 E_{\text{sw}} R_\oplus \sin\phi$$

c) Now solve for an equation that gives $L_{\text{plasmapause}}$ as a function of ϕ . You do this by plugging in the potential you got in part (a) into the equation you obtain in part (b).

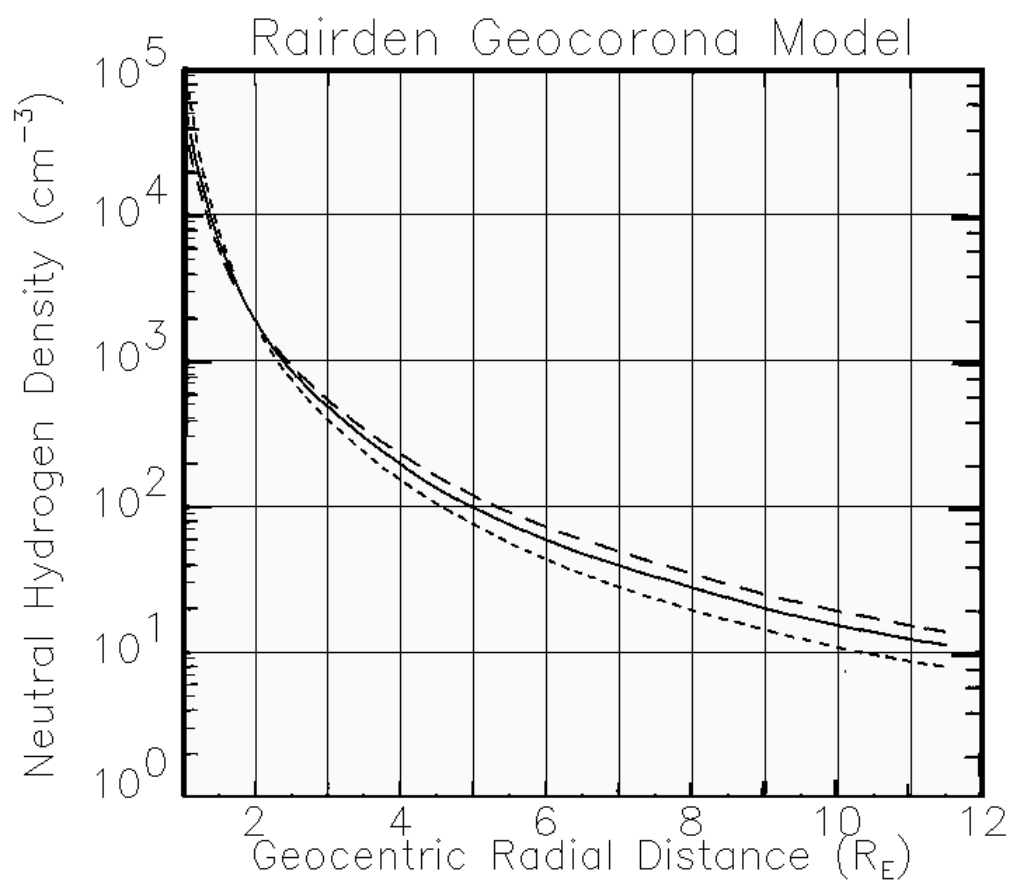
d) Now plot $L_{\text{plasmapause}} = L(\phi)$, with E_{sw} as a separate parameter. ($E_{\text{sw}} = 0.1, 0.3$, and 1.0 mV/m). A polar plot is the best way to present the results. Which produce the result in most obvious agreement with expected locations for the plasmapause?

11. Calculate the energies for which the corotation electric field induced drift velocity $\bar{U} = \omega r \hat{\phi}$, equals the ∇B drift velocity, for protons at $L = 3$, and $L = 6.6$. For simplicity, work in the equatorial plane, and look only at the grad-B drift.

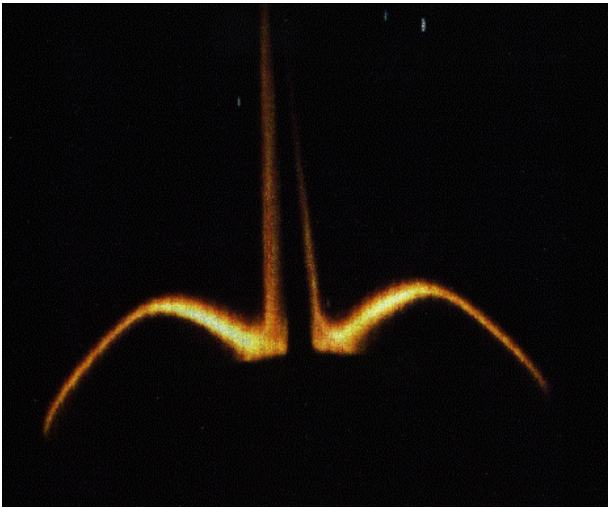
Note:

$$\bar{v}_D = \frac{K_\perp}{qB} \frac{\bar{B} \times \nabla B}{B^2}; \quad \nabla B = -\frac{3}{r} B \hat{r} (\lambda = 0); \quad \left. \frac{\bar{B} \times \nabla B}{B^2} \right|_{\lambda=0} = -\frac{3}{r} -\hat{\theta} \times \hat{r} = -\frac{3}{r} \hat{\phi}$$

$$\bar{v}_D = -\frac{3 K_\perp}{r qB} \hat{\phi}$$



Homework Problem 5.4



Shuttle glow, due to atomic oxygen in the upper atmosphere, came as a rude surprise to astronomers and others. Seen here on STS-39 (right) to be roughly as bright as the aurora.



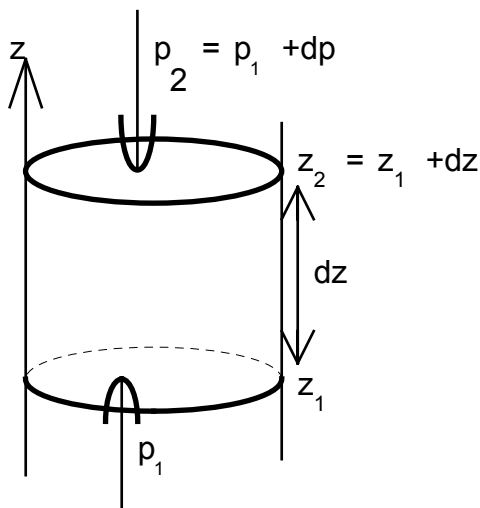
Chapter 6 The Atmosphere

A Introduction

The atmosphere of the earth is important for our purposes as a portion of the satellite (or rocket) environment, and as a source of material for the regions of space outside the sensible atmosphere. The region of interest for our purposes starts at an altitude of about 30 km and extends outward to several thousand kilometers. The neutral atmosphere overlaps the ionosphere, plasmasphere and the radiation belts. That portion of the upper atmosphere which is ionized is termed the ionosphere, extending from about 50 km altitude to ~1000 km altitude. The upper atmosphere is a dynamic system under the influence of gravity, radiation and a complex system of photochemical reactions.

B Law of Atmospheres: Pressure and Density

Our analytic approach to the atmosphere and ionosphere begins with an application of the scale height concept, originally presented in Chapter 1.1



Consider a gas column of unit cross-section as shown, and a coordinate system with z positive upward. If we now consider a slab of gas of thickness dz we can write down the condition that this volume of gas be in equilibrium:

$$\sum \text{Forces}_z = 0 \Rightarrow p_1 = p_2 + \rho g dz \quad (\text{Eqn. 6.1})$$

where

ρ = the mass density of the gas (kg/m^3),
 g = acceleration due to gravity,
 p_1 and p_2 are the pressures,

Figure 6.1

Noting also that $p_1 = p_2 + dp$

. We can combine the above equations into the form

$$p_2 - p_1 = dp = -\rho g dz \quad (\text{Eqn. 6.2})$$

which gives the relation between height and pressure change in differential form. Before we can integrate this equation we must express ρ as a function of p . The ideal gas law can be invoked now. The most useful form for our purposes is:

$$p = nkT \quad (\text{Eqn. 6.3})$$

where

$$\begin{aligned} p &= \text{pressure (N/m}^2\text{)} \\ n &= \text{number density (molecules/m}^3\text{)} \\ k &= \text{the Boltzmann constant} = 1.38 \times 10^{-23} \text{ (J / K)} \\ T &= \text{Absolute Temperature (K)} \end{aligned}$$

The mass density, ρ , is not the commonly used variable in atmospheric work, however. Instead, the mean molecular weight, μ , is used, along with the proton mass, m_p .

$$\rho = n m_p \mu \quad (\text{Eqn. 6.4})$$

$$\begin{aligned} dp &= -n m_p \mu g dz = -\frac{p}{kT} m_p \mu g dz \\ \frac{dp}{p} &= -\frac{\mu m_p g}{kT} dz \end{aligned} \quad (\text{Eqn. 6.5})$$

The quantity $\frac{kT}{\mu m_p g}$ appears to have the dimensions of a height, and is indeed is defined as the pressure scale height:

$$H_p = \frac{kT}{\mu m_p g} \quad (\text{Eqn. 6.6})$$

We can now integrate both sides from some reference level z_r (pressure p_r) to the final values z , and p .

$$\begin{aligned} \int_{p_r}^p \frac{dp}{p} &= -\int_{z_r}^z \frac{dz}{H_p} \Rightarrow \ln \frac{p}{p_r} = -\int_{z_r}^z \frac{dz}{H_p} \\ p(z) &= p_r e^{-\int_{z_r}^z \frac{dz}{H_p}} \end{aligned} \quad (\text{Eqn. 6.7})$$

This is the famous law of the atmospheres. If the scale height is constant with altitude, the integral is trivial:

$$\ln \frac{p}{p_r} = -\frac{z - z_r}{H_p} \Rightarrow p = p_r e^{-\frac{z - z_r}{H_p}} \quad (\text{Eqn. 6.8})$$

For a constant scale height atmosphere we shall now show that H_p is the thickness of this same atmosphere when compressed to a uniform pressure p_r . For an isothermal atmosphere the particle density is proportional to pressure because $p = nkT$, and we can write

$$dn(z) = n_r e^{-z/H_p} dz \quad (\text{Eqn. 6.9})$$

where n is the number density (#/volume) of molecules, and the total number in a 1 m² column is

$$\begin{aligned} n_T &= \int_0^\infty n_r e^{-z/H_p} dz = n_r H_p \int_0^\infty e^{-u} du \\ n_T &= -n_r H_p (0 - 1) = n_r H_p \end{aligned} \quad (\text{Eqn. 6.10})$$

which says that an atmosphere of constant particle density n_r and height H_p will have the same number of particles as the actual column has.

In the actual atmosphere the scale height is approximately constant up to about 120 km altitude at a value of

$$H_p \cong 8.4 \text{ km}$$

As we shall see later, the temperature begins to rise rapidly above 120 km. At the same time the mean molecular mass decreases as shown in Figure 6.2 below. The gravity also decreases slowly (Figure 6.3)

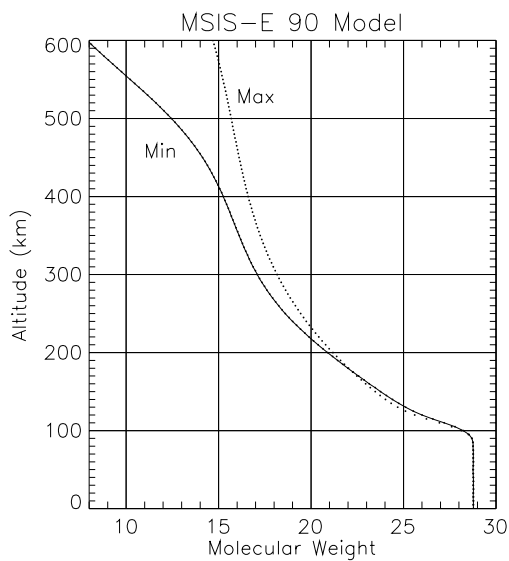


Figure 6.2 Mean molecular weight μ as a function of geometric altitude. MSIS-E Model 1990. Solar minimum (Min) taken 9/22/95; Solar maximum (Max) taken 9/22/89; both at 0° latitude, 20° longitude, local noon.

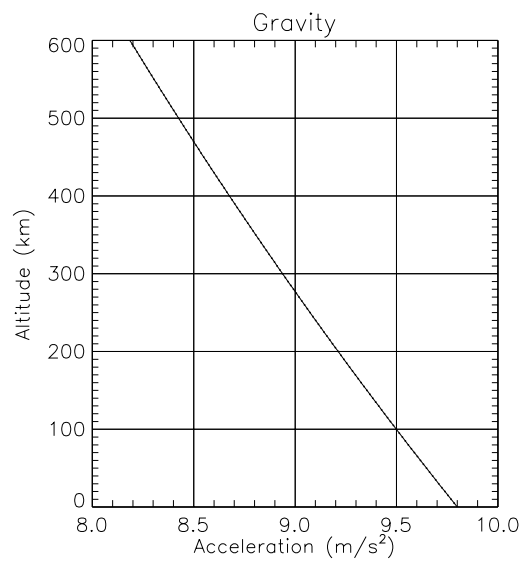


Figure 6.3 Acceleration due to gravity, g , as a function of geometric altitude. (Just a little physics.)

<http://nssdc.gsfc.nasa.gov/space/model/atmos/msis.html>

When these variations are combined with the previously discussed variation of T with altitude we obtain the variation of scale height H_p with altitude. Figure 6.4 shows how the pressure scale height varies with altitude. Since the primary dependence is on temperature, this figure closely resembles the atmospheric temperature profile. Once we have the scale height as a function of altitude we can get the particle density (and the pressure) by integrating the Law of Atmospheres from sea level to any height. Figures 6.5a and 6.5b show how the pressure and density behave for the (model) standard atmosphere. We see from these graphs that both the pressure and the particle density drop off at a slower rate above 100 km which corresponds to the larger scale height at these altitudes, as shown in Figure 6.4

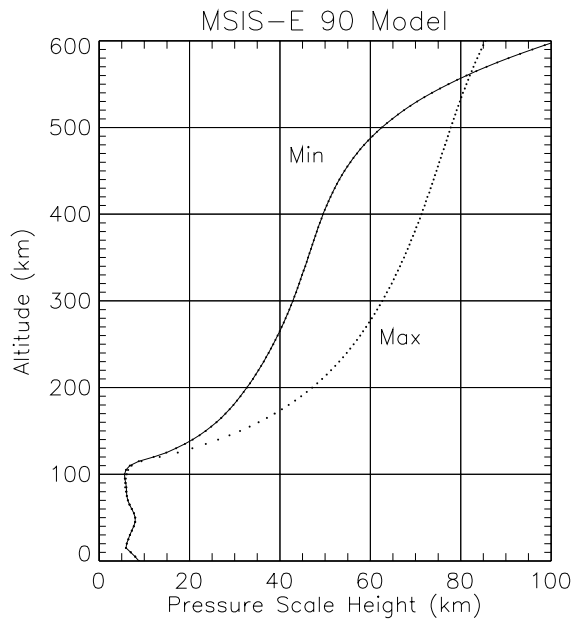


Figure 6.4 Pressure scale height, H_p as a function of geometric altitude. From the NASA MSIS-E-90 Model, implemented at the NASA/GSFC National Space Science Data Center. Calculations are shown for solar minimum (Min) and solar maximum (Max).

<http://nssdc.gsfc.nasa.gov/space/model/atmos/msis.html>

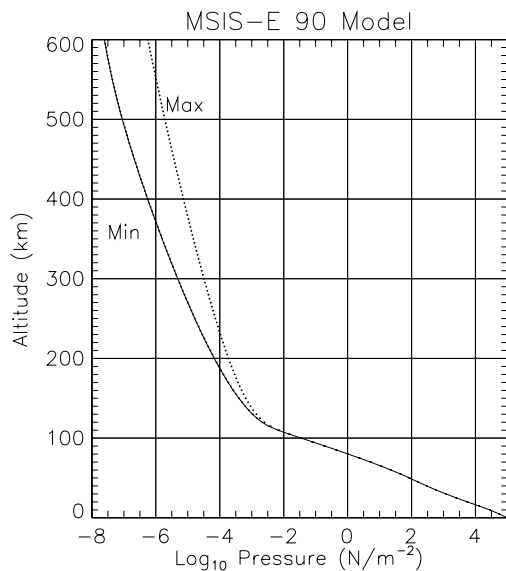


Figure 6.5a Pressure as a function of altitude in the standard atmosphere. Note the log scale on the horizontal axis.

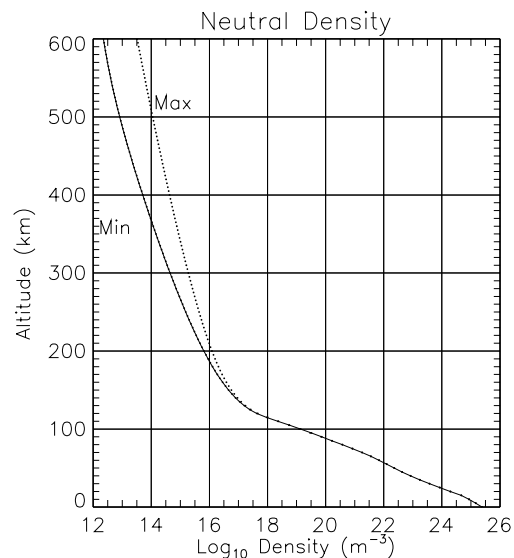


Figure 6.5b Number density as a function of geometric altitude, solar minimum (Min) and solar maximum (Max).

C Temperature of the Atmosphere

The atmosphere is divided up into distinct regions, primarily on the basis of how the temperature varies in different altitude ranges. The different regions are, in order of ascending altitude: the troposphere, the stratosphere, the mesosphere, and the thermosphere. Each of the four characteristic layers, is bounded by a 'pause', a transition region where the temperature profile has an inflection point (reversal in slope). (see Figure 6.6). The temperature gradient is called a 'lapse rate'.

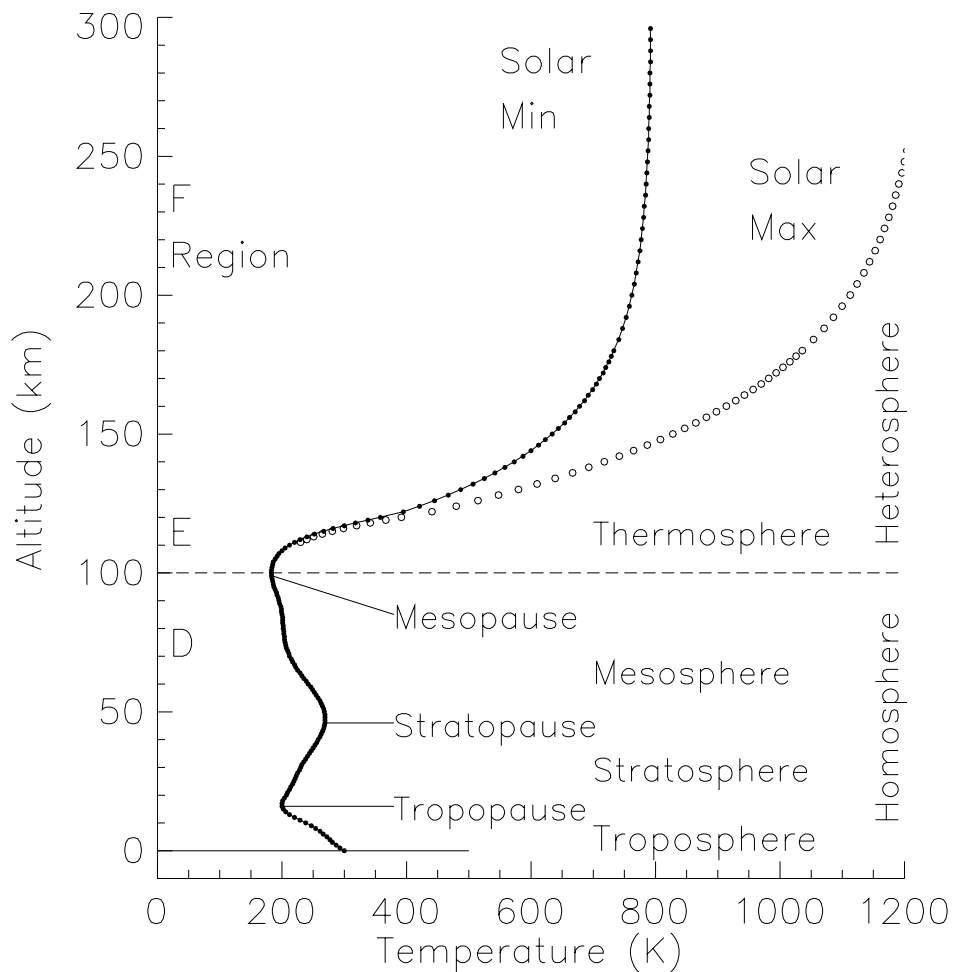


Figure 6.6 Variation in temperature with altitude. Figure created with MSIS-E 1990 model. Ionospheric regions (D, E, F, annotated along left hand axis).

The lowest region in the atmosphere is called the troposphere where clouds and weather originate and where turbulence is the dominant physical process. The temperature decreases at a fairly uniform rate of -6.5°C/km for about 15 km and reaches a minimum of approximately -60°C (for mid latitudes) at the tropopause. If we calculate a convective temperature lapse rate for a dry atmosphere under adiabatic conditions we get a value of about -9°C/km . The difference is mostly due to the presence of water vapor which modifies the purely convective cooling in two ways (a) The condensation of water vapor to form clouds gives up energy which heats the air and (b) water molecules are excellent absorbers (and radiators) of IR radiation emitted by the surface of the planet, a process which leads to further heating of the air (the so-called Greenhouse Effect). The net result of these processes is to decrease the temperature lapse rate to the observed value of -6.5°C/km . The tropopause is where the temperature gradient vanishes. It occurs at an altitude of 6 km to 18 km (higher and colder over the equator). It is the domain of high winds and the highest observed cirrus clouds.

The next region is the stratosphere which is about 25 km thick at mid latitude (thicker over the poles and thinner over the equator). The temperature in the stratosphere increases with increasing altitude as a result of the absorption of solar ultraviolet radiation near the base of the stratosphere. This effect is particularly pronounced in the ozone layer where a concentration of only a few parts per million of O_3 leads to very pronounced absorption of wavelengths in the 200-300 nm region and subsequent heating of the atmospheric gases. The temperature increase terminates at the stratopause which occurs at an altitude of about 50 km and is followed by another region of declining temperature, the so-called mesosphere. The temperature in this region is mostly determined by convection, the warmer stratosphere now acting in much the same way as the planetary surface acts with respect to the troposphere. This cooling trend terminates at the mesopause where the coldest temperatures of the atmosphere (about 180°K) are observed.

Above about 80 km we enter the thermosphere where strong positive temperature gradients exist, caused by a combination of facts: (a) the density of the gases is becoming quite low (b) there are few if any triatomic molecules present (c) solar radiation of short wavelength ($\lambda < 175\text{ nm}$) heats the atmosphere from above (d) exothermic reactions between ions and electrons occur.

Since the heating now occurs from above rather than below there is no convection and the temperature increases although we are now entering a region where collisions are too infrequent to establish thermodynamic equilibrium. At extreme altitudes (300 - 400 km) the temperature rise stops as we enter the exosphere where the temperature becomes essentially independent of altitude.

The high temperatures found in the upper reaches of our atmosphere do not imply that there is a large heat reservoir there, since particle densities are very low at these altitudes. Instead these high temperature (approaching $1000 - 1500^{\circ}\text{K}$) merely reflect the inefficiency of the heat removal processes. Only about 10^{-6} of the solar energy supplied to the earth is absorbed in the thermosphere, the rest being absorbed in the lower atmosphere and the ground.

There also exists considerable variability of this terminal temperature both as function of solar activity (900° to 1700°) and on a daily basis. These variations result from the low heat capacity of the upper atmosphere so that fluctuations in the UV radiation are rapidly reflected in temperature changes. Geomagnetic storms can also result in large temperature fluctuations of the upper atmosphere probably due to increased particle precipitation.

D Composition of the Atmosphere

If we divide the atmosphere according to composition there are three characteristic spheres:

(a) Homosphere: the region of essentially uniform composition in the sense that the mean molecular weight remains constant as we proceed from the surface upward. Even as we traverse the region of maximum ozone concentration we find that the mean molecular weight remains essentially constant.

(b) Heterosphere: At about 100 km we enter the region of significantly varying composition. The "molecular weight" diminishes from 29 at about 90 km to about 16 at 500 km. Above the level of oxygen dissociation, nitrogen begins to dissociate and diffusive separation sets in.

(c) Exosphere is the region in which the escape of molecules from the atmosphere becomes significant. The lower limit of the exosphere is not clearly defined since the escape depends on the mass of the atoms involved, but values of 500 km to 1000 km are typical.

Table 6.1. Normal Composition of Clean, Dry Atmospheric Air near Sea Level^a

Constituent gas and formula	Content, percent by volume	Molecular weight ^b
Nitrogen (N ₂)	78.084	28.0134
Oxygen (O ₂)	20.9476	31.9988
Argon (Ar)	0.934	39.948
Carbon Dioxide (CO ₂)	0.0314 ξ	44.00995
Neon (Ne)	0.001818	20.183
Helium (He)	0.000524	4.0026
Krypton (Kr)	0.000114	83.80
Xenon (Xe)	0.0000087	131.30
Hydrogen (H ₂)	0.00005	2.01594
Methane (CH ₄)	0.0002 ξ	16.04303
Nitrous oxide (N ₂ O)	0.00005	44.0128
Ozone (O ₃)	Summer: 0 to 0.000007 ξ	47.99827
	Winter: 0 to 0.000002 ξ	
Sulfur Dioxide (SO ₂)	0 to 0.0001 ξ	64.0628
Nitrogen Dioxide (NO ₂)	0 to 0.000002 ξ	46.0055
Ammonia (NH ₃)	0 to trace	17.0306
Carbon Monoxide (CO)	0 to trace	28.01055
Iodine (I ₂)	0 to trace	253.8088

^aAdapted from Dubin, Sissenwine, and Wexler [1].

^bOn basis of carbon-12 isotope scale for which C¹² = 12.

ξThe content of these gases may under go significant variations from time to time or from place to place *relative* to the normal indicated for those gases.

The composition of the neutral atmosphere changes with increasing altitude as a result of chemical (including photochemical) reactions. The region of the atmosphere in which chemical reactions are prevalent is called the chemosphere and extends from roughly 20 km to 110 km, although some reactions occur both above and below this region. Table 6.1 shows the normal composition of the atmosphere near sea level.

Apart from the variable water vapor content the composition of the atmosphere is approximately constant up to an altitude of 30 km. At this altitude the concentration of ozone (O_3) begins to increase and because of its very high absorption coefficient for ultraviolet radiation it plays a dominant role in the physical processes of the atmosphere.

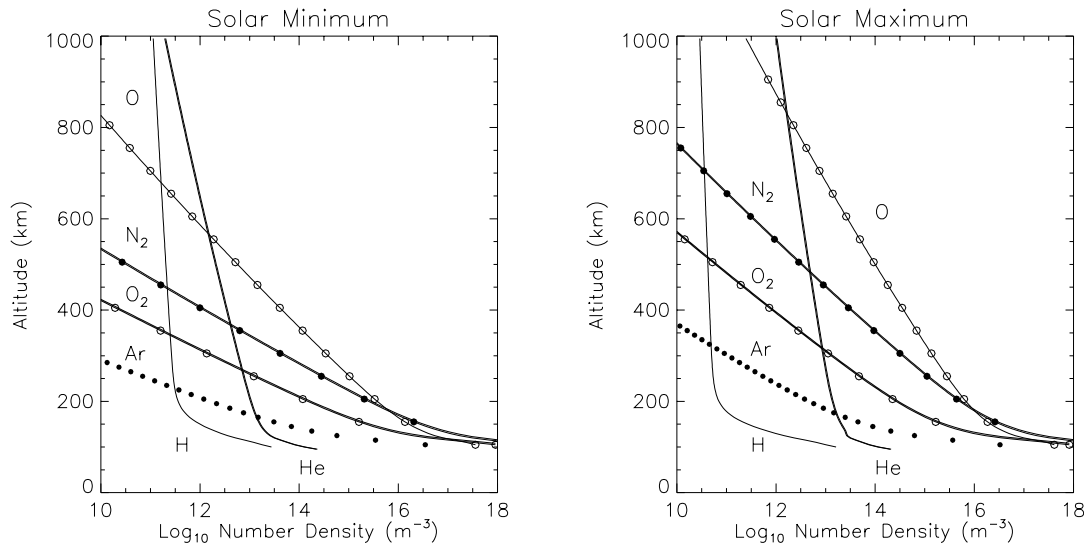


Figure 6.7 MSIS-E 1990 model. Relative concentrations of atmospheric constituents.

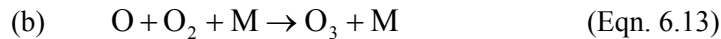
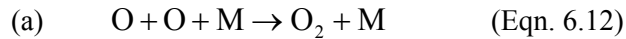
As the altitude increases further the ozone concentration decreases and at about 100 km atomic oxygen becomes an important constituent. Nitrogen remains mostly in its diatomic form and hence the region between 100 to 1000 km is a mixture of molecular oxygen, atomic oxygen, and molecular nitrogen with atomic oxygen increasing in relative concentration as the altitude increases. Between 1000 km and 2000 km helium appears to be the most important constituent and above 2000 km the atmosphere consists essentially of hydrogen.

The major photochemical process occurring in the atmosphere is the dissociation of molecular oxygen



which can occur at wavelengths below 250 nm and represents a very efficient absorption mechanism for the far ultraviolet portion of the solar spectrum.

One important loss mechanism for atomic oxygen is the three body recombination reaction which can take two forms:



where the third body M takes up some of the energy and momentum of chemical bond formation. Since these two reactions are but two of literally dozens which enter in the establishment of the steady state it is obvious that a quantitative determination of actual equilibrium concentrations of say O_3 as a function of altitude and time is a very complicated process involving reaction rates for many processes. Nevertheless, we can roughly see how the ozone layer represents a steady state determined by production of O_3 :



The two major loss mechanisms are:

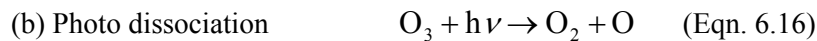
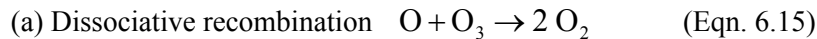


Photo dissociation of ozone occurs very efficiently for photons in the 200 to 300 nm wavelengths region. Although the ozone concentration never exceeds a few parts per million even in the so-called "ozone layer" it nevertheless represents a most effective filter for potentially harmful uv radiation. At 255 nm where the ozone absorption curve peaks the absorption coefficient is about 276 cm^{-1} , the ozone layer has a thickness of about 0.25 cm (at $p = 760 \text{ mm Hg}$ and $T = 18^\circ \text{ C}$).

$$\text{Thus the vertical attenuation is } I = I_0 e^{-0.25 \times 276} \quad \text{or} \quad \frac{I}{I_0} = e^{-69} \approx 10^{-30}$$

which is a pretty good attenuation considering how few ozone molecules there are in this so called ozone layer.

E Transmission of Electromagnetic Waves Through the Atmosphere

As shown in Figure 6.8 the atmosphere is largely opaque to EM waves but there are several important "windows" through which radiation of certain wavelengths does reach the surface of the earth. Practically all our knowledge of the external universe comes to us through these "windows". Furthermore the photon energy of the radiation passing through the "optical window" is just right to sustain biological processes. If that window were to shift in either direction by an appreciable amount life as we know it would not be possible. We shall consider 8 regions of the spectrum (Figure 6.8) and briefly discuss the various processes in the atmosphere which do (or don't) absorb the incident radiation.

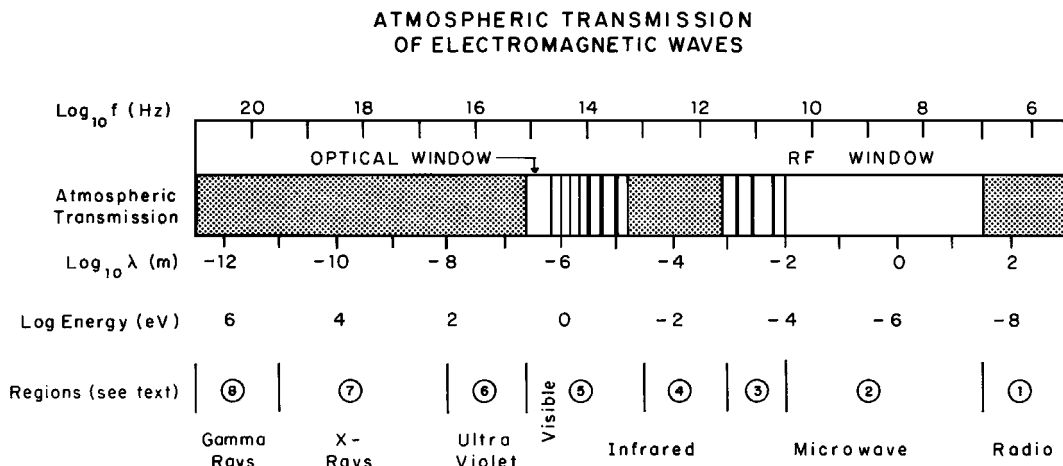


Figure 6.8 Atmospheric transmission of electromagnetic waves

1) RF Region:

Strong interaction (not necessarily absorption) up to 5 or 10 MHz. Any frequency below the plasma frequency f_p ($f_p = 9\sqrt{n_e}$) will not propagate, hence the atmosphere is opaque (see chapter 7)

2) RF Window:

10 MHz - 20 GHz.

No effective absorption process occur, hence the atmosphere is transparent. Frequency region of radio astronomy.

3) Microwave Absorption Region:

20 - 300 GHz ($\lambda = 1.5$ cm to $\lambda = 1$ mm). Absorption by molecular rotations. H_2O and O_2 have major rotational modes at 22 and 60 GHz respectively. Absorption in this region is intermittent with respect to wavelength.

4) Infrared Absorption Region:

$\lambda = 1000$ to 30 μm ($f = 3 \times 10^{11}$ to 10^{13} Hz). Absorption by vibrational transitions in H_2O , CO_2 and to some extent O_3 . Rotational levels spread vibrational levels into bands. The Black Body spectrum of the earth is $4 - 60$ μm peaking at 15 μm . Thus the long wavelength part of the earth radiation is completely absorbed by the atmosphere.

5) Optical Window: $\lambda = 30 \mu\text{m}$ to $.3 \mu\text{m}$.

In the IR portion of the window ($30\mu\text{m}$ to $.7\mu\text{m}$) there is intermittent absorption by all molecular species present (H_2O , CO_2 , N_2O , CH_4 , O_3 , etc.) leading to strongly frequency dependent transmission. In the visible portion of the window ($0.7 \mu\text{m}$ to $0.4 \mu\text{m}$) there is very little absorption.

6) UV and Soft X Ray Region: $\lambda = 300$ to 10 nm . ($3000 \text{ \AA} - 100 \text{ \AA}$)

Near UV Region: Absorption by photo dissociation of O_3 . At shorter wavelengths electronic transitions in O_2 . Below 100 nm absorption by electronic transitions in N_2 and O .

7) X-Ray Region: $\lambda = 10$ to $.01 \text{ nm}$. ($100 \text{ \AA} - 0.1 \text{ \AA}$)

Absorption by excitation of inner electrons in both atomic and molecular species (mostly N_2 and O_2).

8) Gamma Ray Region: $\lambda < .01 \text{ nm}$.

Absorption by Photoelectric Effect and Compton Scattering in all species present in atmosphere.

Some perspective on the absorption and transmission windows can be obtained by means of the curve shown in Figure 6.9. This figure gives the depth of penetration for various wavelengths into the atmosphere from above. The atoms and molecules responsible for the absorption are identified, and in particular the fate of Lyman α and Lyman β .

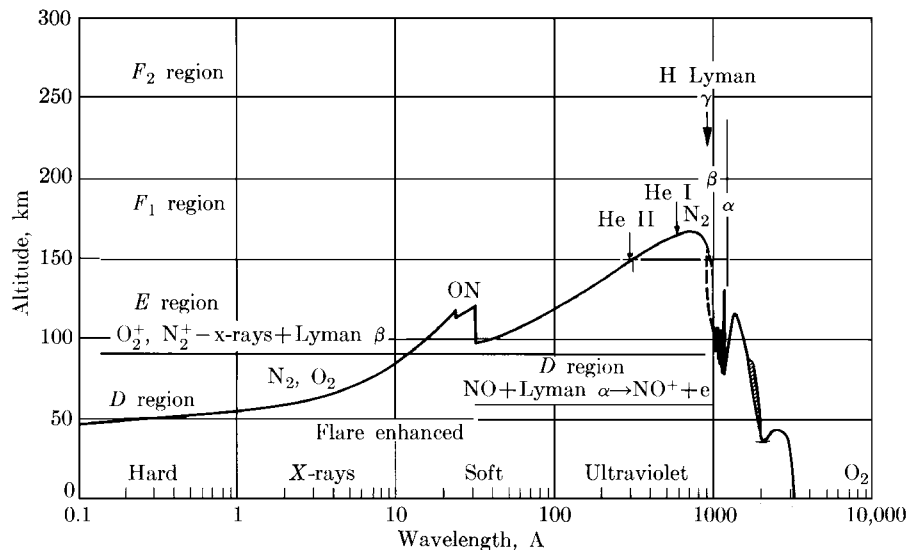


Figure 6.9 The depth to which $1/e$ of the incident solar intensity penetrates, for 0.1 to $10,000 \text{ \AA}$. Copied from "Atomic and Space Physics", p 131, by Alex Green and Philip Wyatt, 1965. ($1 \text{ nm} = 10 \text{ \AA}$)

Originally: R. Tousey, "Ultraviolet Spectroscopy of the Sun", Chapter 1, Space Astrophysics, W. Liller ed., 1961

Finally perspective on the visible and IR portions of the electromagnetic spectrum is given in Figure 6.10, which shows characteristic black body spectra for the sun and earth, along with the absorption percentage on the ground, and at 11 km altitude. Note the considerably diminished effect of H₂O at 11 km altitude.

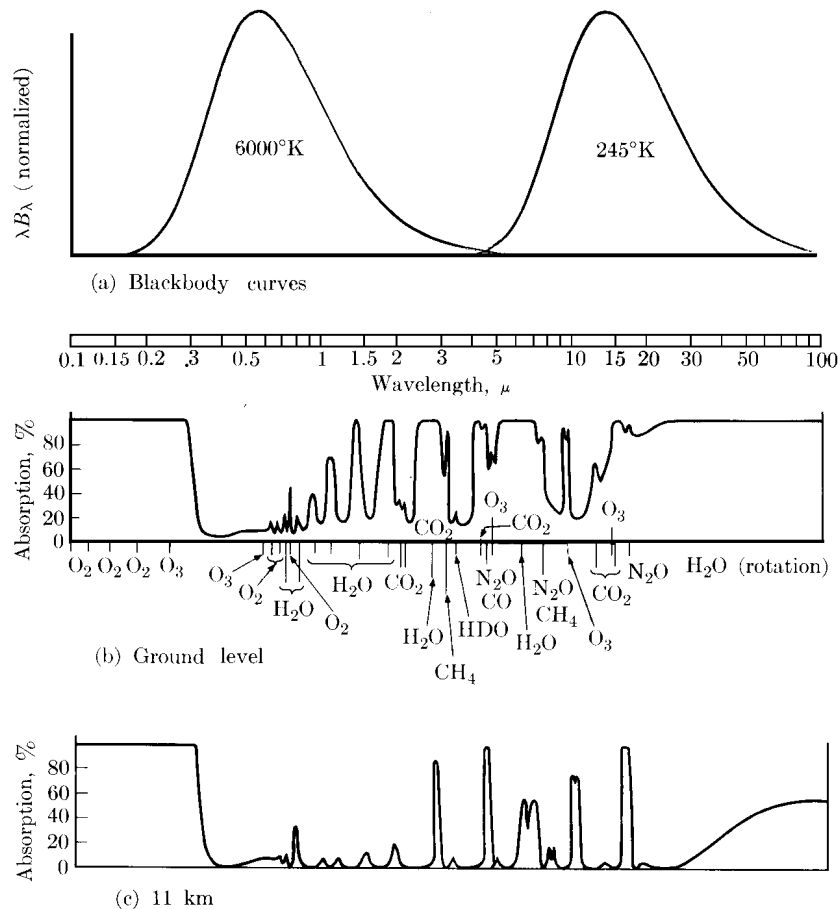


Figure 6.10. Atmospheric absorptions. (a) Blackbody curves for 6000 K and 245 K. (b) Atmospheric gaseous absorption spectrum for a solar beam reaching ground level. (c) The same for a beam reaching the temperate tropopause. The axes are chosen so that areas are (a) proportional to radiant energy. Integrated over the earth's surface and over all solid angles the solar and terrestrial fluxes are equal; consequently, the two blackbody curves are drawn with equal areas beneath them. An absorption continuum has been drawn beneath bands in (b). This partly hypothetical because it is difficult to distinguish from the scattering continuum, particularly in the visible and near infrared spectrum. Conditions are typical of midlatitudes and for a solar elevation of 40° or diffuse terrestrial radiation. Copied from "Atomic and Space Physics", p 130, by Alex Green and Philip Wyatt, 1965. Originally from: R. M. Goody, Atmospheric Radiation I. Theoretical Basis, Clarendon Press, Oxford, 1964.

F Problems

1. Estimate the scale height above 200 km altitude, using the information in Figure 6.5b. How does this compare to the values in Figure 6.4? Now calculate the scale height for N_2 using a temperature of 1000 K (ignore the altitude dependence of g).
2. If the temperature of the upper atmosphere doubled (from 1000 to 2000 K), how would the number density at 700 km altitude change (give a number). In order to do this you should use the scale height calculated in problem 1, assume the atmosphere is isothermal, and that the density at 200 km does not change. (This type of temperature increase is the source of some of the changes seen in Figure 6.7)
3. A) If the temperature of human skin is 98.6 F, at what wavelength does the blackbody radiation from skin peak? Which IR window is best for night time surveillance of humans?
B) If the average temperature of the stacks on a surface ship is 750 K, at what wavelength does blackbody radiation from the stacks peak? Which IR window is best for night time surveillance of ships?



Northern Lights by Fridtjof Nansen (1910-1911) - Cover illustration from his book *Nord I Tækeheimen* (In Northern Mists) The image of the northern lights is from a woodcut created from a crayon sketch he made in 1893.

Chapter 7 The Ionosphere

A Introduction

The upper atmosphere, beginning at about 50 km altitude, is partially ionized by ultraviolet and x-ray radiation from the sun. This region of partially ionized gas extends upwards to about 1000 km altitude. The region is termed the ionosphere. The ionosphere is important as a source of plasma for the magnetosphere, and as a medium which reflects radio waves at frequencies from a few Hz up to several Megahertz. Like the rest of the earth-sun system we have explored it is a dynamic region with an amazing variety of features. The most striking is the aurora, which reflects the influence of the hot magnetospheric plasmas on the high latitude atmosphere. Figure 7.1 shows an auroral image acquired by DMSP, in a snapshot of the auroral activity found in visible wavelengths. (DMSP orbits at an altitude of about 500 km, looking down on the atmosphere. The aurora occurs at altitudes below 100 km.) These bright glows are visible from the surface of the earth, and the northern lights (aurora borealis) and southern lights are a topic of perpetual interest for space physicists.

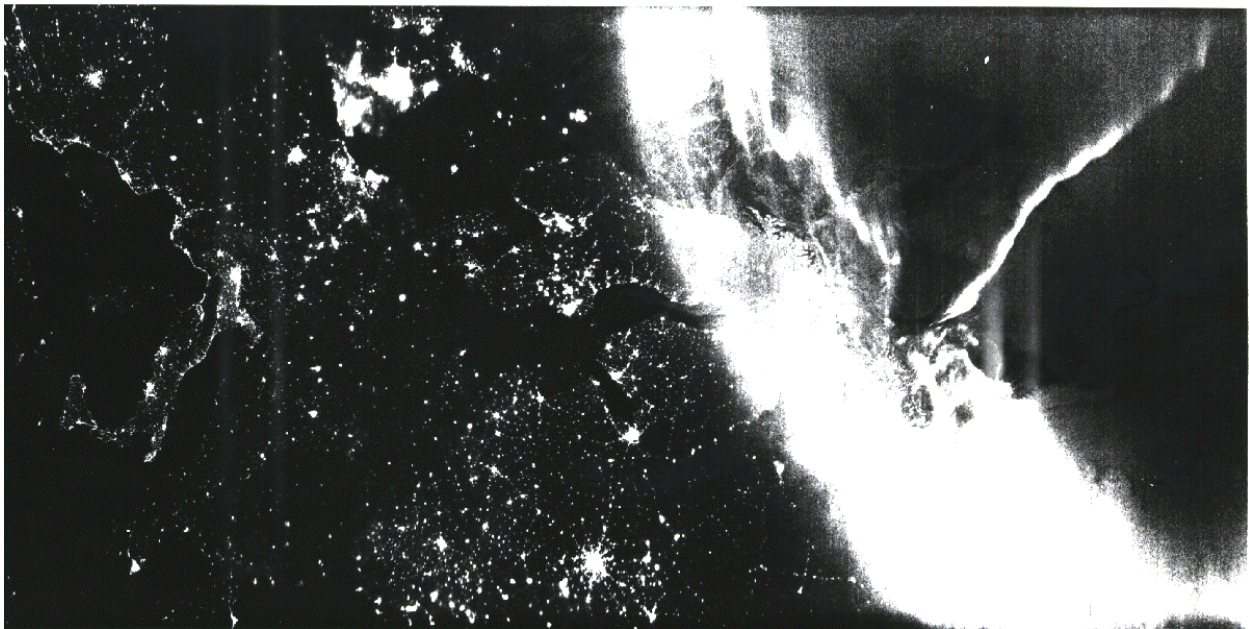


Figure 7.1 Auroral image from the Defense Meteorological Satellite Program (DMSP) Operational Linescan System. The OLS is the visible imager, with 2.7 km resolution (low resolution mode). This image of the aurora was obtained over Europe on 11 December 1988, at ~2120 UT. The satellite traverses a polar orbit from 76.8 N, 45.8 E (far right) to 39.6 N, 9.7 E (far left) over a 10 minute interval. The 'boot' of Italy is on the left, the bright spot on the center, lower edge of the image is Moscow. The auroral oval cuts across the northern edge of Scandinavia in this image, extending 8-10 degrees in latitude at a given longitude. There is a small 'polar' arc extending northward from the main auroral glow. Image courtesy of Dr. Charles Pike, USAF (Geophysics Lab, Hanscom AFB) See also <http://web.ngdc.noaa.gov/dmsp/source/aurora.html>.

Embedded in the upper atmosphere there are several fairly distinct layers where positive ions and electrons are present in sufficient numbers to affect radiowave propagation. The variations in

electron density with altitude led to the subdivision of the ionosphere into the D, E and F regions as shown in Fig. 7.2. The electron density in all of these regions varies with time of day, season and solar activity. On the average, the electron density is greater in the daytime than at night, greater in summer than in winter, and greater during the maximum phase of the sunspot cycle than at sunspot minimum. However, there are noteworthy exceptions. For instance, the peak electron density of the F₂-layer is greater in winter than in summer. These exceptions will be described in section 7.

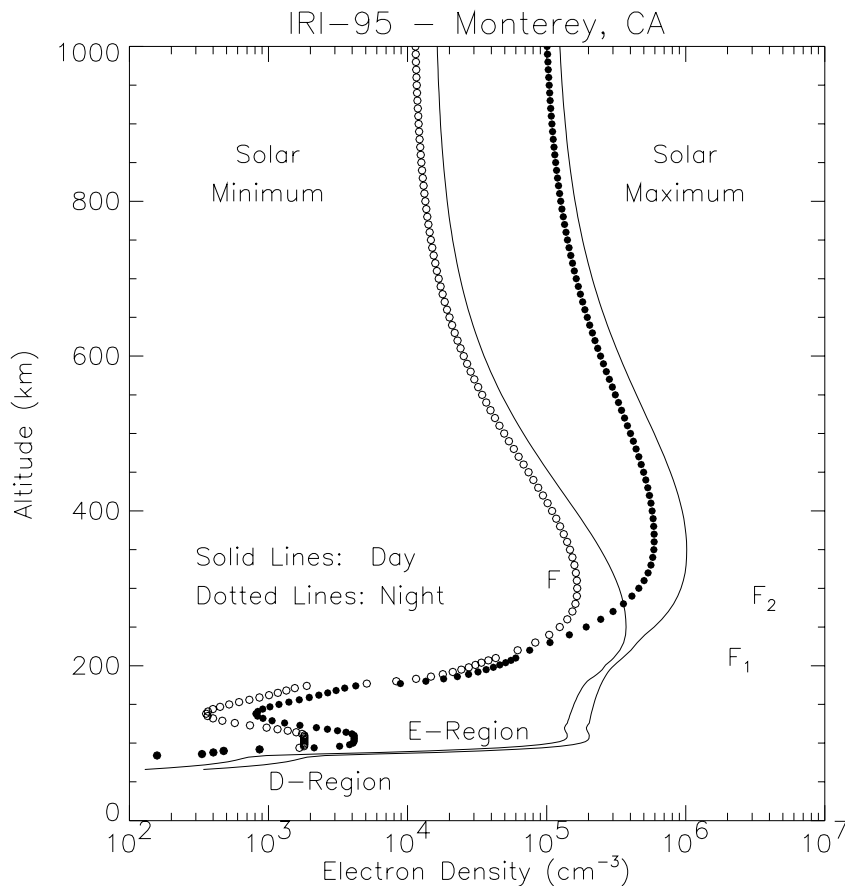


Figure 7.2 Electron density as a function of altitude for the mid-latitude ionosphere. The International Reference Ionosphere 1995 (IRI-95) model was run for the location of Monterey, CA (geographic latitude = 36.5°, geographic longitude = 238°), for July 4, 1989 (solar maximum) and July 4, 1995 (solar minimum). Calculations were done at 0LT (local midnight), and 12 LT (local noon). Note that for the more traditional metric unit for density, m⁻³, you multiply by 10⁶. [<http://nssdc.gsfc.nasa.gov/space/models/iri.html>]

The ionosphere serves as a high altitude reflector for short-wave broadcasting and long-range communication. High frequency (3-30 MHz) radio waves are reflected by the F-region or the daytime E-region, while very-low-frequency (less than 30 kHz) radio waves are reflected by the D-region. When a radio wave travels through the ionosphere, the electrons are set into oscillation at the frequency of the wave. The energy of oscillation is obtained from the wave. The electrons lose some of this energy by colliding with the neutral molecules of the upper atmosphere. In these collisions, much of the energy of oscillation that the electrons acquired from the radio wave is transferred to the neutral molecules, and appears as random kinetic energy. Consequently, the radio wave is attenuated. The higher the rate of collision of electrons with neutral molecules the

greater is the attenuation. Hence, most of the attenuation generally occurs in the D-region, where the density of the neutral molecules and the collisional frequency of the electrons is greater than in the upper region.

Furthermore, it turns out that the attenuation is inversely proportional to the square of the frequency, so that waves of frequency near the lower end of the HF band are more severely attenuated than those near the upper end of the band. In practice, therefore, the optimum frequency used by a communication circuit is the highest frequency that will be reflected by the ionosphere.

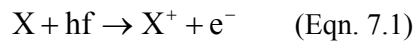
During the occurrence of a solar flare, the enhanced x-ray radiation from the sun causes the electron density in the D-region to increase by a large factor. This results in the complete absorption of high-frequency radio waves, and the disruption of short-wave communication over the earth's sunlit hemisphere. Beginning about 26 hours after a solar flare, the global F₂-region electron density undergoes substantial variations owing to the interaction of the disturbed solar wind with the ionosphere. These variations could last for several days after the flare.

At high latitudes, a significant fraction of the ionization is produced by charged particles that have been dumped from the radiation belts at times of magnetic storms, and also by charged particles from the sun that have been diverted to these latitudes by the geomagnetic field.

B Formation of an Ionospheric Layer (Chapman Theory)

The ionospheric plasma is electrically neutral i.e. there are an equal number of positive and negative charges in a reasonable size volume. But due to their much larger mass (typically a factor of 30,000 or more) the ions are much less mobile and their motions can be ignored for many purposes such as the interaction with radiowaves. The ionosphere is not entirely or even predominantly composed of charged particles. Neutral particles far outnumber the ions and electrons even in the layers of highest charge concentration. For example in the F₂ region the charged particle density is only about 10^{-3} that of the neutral particle density.

The primary process by which charged particles are produced in the upper atmosphere is the ionization of the neutral gases by solar ultraviolet and x-radiation, particularly at low and mid latitudes. The formation of layers in the ionosphere can be qualitatively understood by considering the ionization of a typical atom or molecule in the atmosphere



The rate of this reaction, $q(z)$, at a given altitude, z , will be proportional to the product of the concentration of the molecules, $n(z)$, and the intensity of the solar radiation, $I(z)$, at that altitude. Thus,

$$q(z) \propto n(z) I(z) \quad (\text{Eqn. 7.2})$$

At very high altitudes, there are lots of photons but few molecules. At very low altitudes there are lots of molecules but very few photons to cause ionization. Thus, the rate of electron production $q(z)$ peaks at some intermediate level as illustrated in Figure 7.3

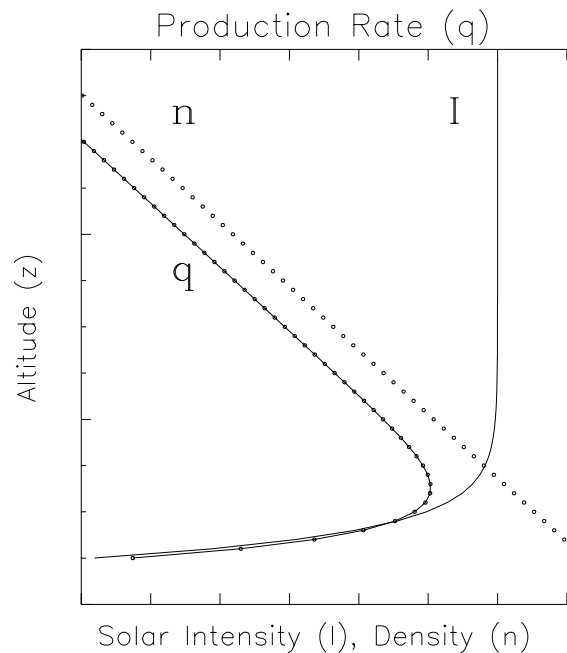


Figure 7.3 Schematic illustration of Chapman's model

In 1931, Sidney Chapman worked out a simplified theoretical model of this process by assuming that:

the atmosphere consists of only one gas, and is horizontally stratified so that the concentration, n , of atoms (or molecules) depends only on the altitude, z ;

the atmospheric gas density decreases exponentially with altitude, as $n(z) = n_0 e^{-z/H}$, and;

(3) the radiation from the sun is monochromatic.

Despite the rather drastic simplifications the theory yields remarkably good results for each atmospheric constituent. The conclusions of Chapman's theory may be summarized as follows.

1 Solar Intensity

First, the intensity of the solar radiation at the altitude z is given by:

$$I(z) = I_{\infty} e^{-\sigma n(z) H \sec \chi} \quad \text{photons/m}^2 \text{ s} \quad (\text{Eqn. 7.3})$$

where I_{∞} is the unattenuated intensity, i.e. the intensity at the “top” of the atmosphere, is the **absorption** cross-section of the gas for the particular wavelength of interest ($\sigma \sim 10^{-22} - 10^{-27} \text{ m}^2$ for the earth’s atmosphere at typical wavelengths of solar radiation. $n(z)$ is the number density of the atmospheric gas molecules at the altitude z ,

H is the scale height of the atmosphere,

$\left(\frac{kT}{mg}\right)$ and χ is the angle between the direction of the solar radiation and the vertical; χ is often called the solar zenith angle.

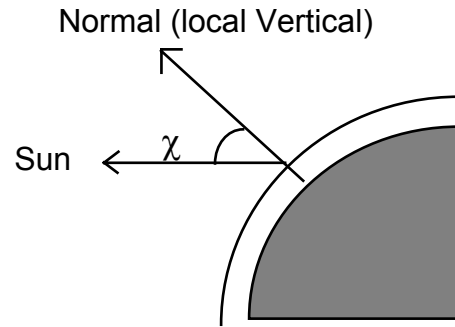


Fig. 7.4 Definition of solar zenith angle

Note that (a) $n(z)$ increases and consequently $I(z)$ decreases as the radiation propagates downward, and (b) $I(z)$ is greater at noon, when χ is small than at sunrise or sunset when χ is large.

2 Peak Altitude

Assuming that each photon produces one electron, the rate of electron production at the altitude z is:

$$q(z) = \sigma n(z) I(z) \quad (\text{Eqn. 7.4})$$

(In reality, the primary electron produced by a highly energetic photon will itself produce several additional electrons by cascaded collisions, in which case the above expression should be modified by the effective number of electrons produced by each photon. This modifying factor can range from 10-20 in the UV range. Note that due to an unfortunate shortage of Greek symbols, the same symbol ‘ σ ’ is used here for the **ionization** cross section. It will be similar to, but not identical, to the absorption coefficient.)

Now, the number density n_m at the altitude where $q(z)$ is maximum is given by :

$$\sigma n_m H \sec \chi = 1 \quad (\text{Eqn. 7.5})$$

Hence, the altitude at which $q(z)$ is a maximum is

$$z_m = H \ln (n_o \sigma H \sec \chi) \quad (\text{Eqn. 7.6})$$

At this altitude, the intensity of the solar radiation is $I_{\infty} e^{-1} = I_{\infty}/e$. (Eqn. 7.7)

Note that z_m is independent of the intensity of the solar radiation. It depends only on the properties of the atmosphere and the solar zenith angle. When χ increases, so does z_m . Thus the altitude of peak electron production is higher at sunrise and sunset than at noon.

example: using $n_0 = 2 \times 10^{25} \text{ m}^{-3}$, $H = 8.71 \times 10^3 \text{ m}$

$$\begin{array}{llll} \chi = 0^\circ & \sigma = 10^{-18} \text{ cm}^2 & (10^{-22} \text{ m}^2) : & h_m = 145 \text{ km} \\ & \sigma = 10^{-20} \text{ cm}^2 & (10^{-24} \text{ m}^2) : & h_m = 105 \text{ km} \end{array}$$

$$\begin{array}{llll} \chi = 60^\circ & \sigma = 10^{-18} \text{ cm}^2 & : & h_m = 151 \text{ km} \\ & \sigma = 10^{-20} \text{ cm}^2 & : & h_m = 111 \text{ km} \end{array}$$

3 Peak Production Rate

The rate of electron production at the peak is given by:

$$q_m = \sigma n_m \frac{I_\infty}{e} = \sigma \left(\frac{1}{\sigma H \sec \chi} \right) \frac{I_\infty}{e} = \left(\frac{I_\infty}{e H} \right) \cos \chi \quad (\text{Eqn. 7.8})$$

[Note: here as above, 'e' is the 'exponential', not the charge of the electron]

Note that q_m is proportional to I_∞ and $\cos \chi$. As χ increases q_m decreases. Thus the peak electron production is greater at noon than at sunrise and sunset.

Example (hypothetical)

$$I_\infty = 5 \times 10^{11} \text{ photons/m}^2 \text{ sec}, \quad H = 8.71 \times 10^3 \text{ meters}, \quad \chi = 10^\circ$$

$$\text{then } q_m = \frac{5 \times 10^{11}}{(2.72)(8.71 \times 10^3)} \cos 10^\circ = 2.08 \times 10^7 \text{ electrons/m}^3 \text{ sec}$$

4 Production Rate

Now, define q_0 to be the peak rate of electron production when $\chi = 0$, and z_0 to be the altitude for the peak production when $\chi = 0$.

$$\text{Then, } q_0 = \frac{I_\infty}{eH} \quad \text{and} \quad q_m = q_0 \cos \chi \quad (\text{Eqn. 7.9})$$

$$\text{Also, } z_0 = H \ln(n_0 \sigma H) \quad (\text{Eqn. 7.10})$$

A careful rearrangement of all the terms eventually leads to:

$$q(h) = q_0 \exp[1 - h - \sec \chi \exp(-h)] \quad (\text{Eqn. 7.11})$$

$$\text{where: } h = \frac{z - z_0}{H}$$

This is called the Chapman (ion-pair) production function. Plots of $q(h)$ versus h for various values of χ are shown in Figure 7.5.

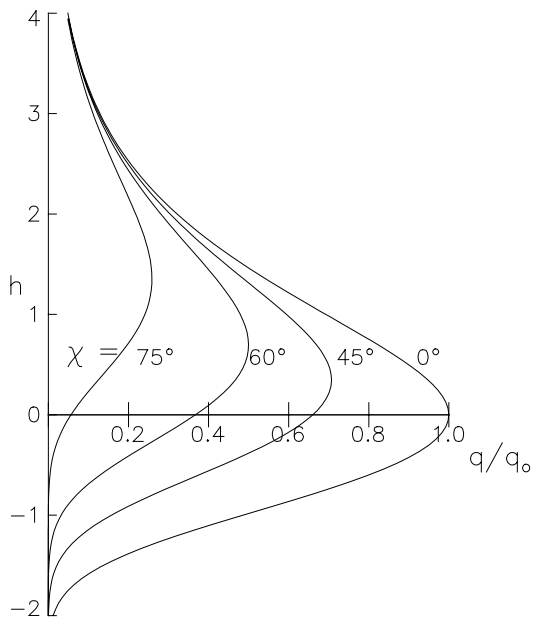


Figure 7.5 : Normalized Chapman production rate versus reduced height, z , parametric in solar zenith angle χ , at the equator

6 Loss Processes

The above discussion of how the ionization process occurs provides the major fraction of the knowledge need to understand why the Chapman layers exist. The ions which are generated in this fashion can decay quite quickly, however, because direct collisions between the ions and neutrals lead to the elimination of the plasma. A primary loss mechanism is dissociative recombination, where electrons combine with various molecular (positive) ions.

C Ionosphere morphology - Altitude Structure

Different atmospheric gases are ionized by various components of the solar radiation to produce the different ionospheric regions.

1 D-region: 60 km to 90 km.

Formation

This is the lowest region of the ionosphere and is thus produced by the penetrating component of the incident radiation, namely short wavelength ultraviolet (Lyman α with $\lambda = 121.6$ nm) and x-rays. The D region is formed primarily by the ionization of the trace atmospheric constituent NO ($[\text{NO}] \approx 10^7 \text{ cm}^{-3}$ as compared with $[\text{N}_2] \approx 10^{14} \text{ cm}^{-3}$ at 85 km) by the relatively intense Lyman- α radiation ($I_\infty \approx 3.3 \times 10^{11} \text{ photons/cm}^2 \text{ sec}$) from the sun. Ionization of N_2 and O_2 by solar x-rays is a secondary process. The contribution of this latter process is small except during a solar flare. The dominant loss process is the dissociative recombination of electrons with various molecular positive ions.

Electron Density

The electron density increases from about $100 \text{ electrons/cm}^3$ at 60 km to about $10^4 \text{ electrons/cm}^3$ at 90 km, around noon. It is greater in the summer than in the winter, and greater at sunspot maximum. At night, when there is no incident radiation, the electrons quickly recombine with the molecular positive ions, so that the D-region disappears, except at latitudes greater than about 65° , where particle bombardment sustains the ionization.

Anomalies

Sudden Ionospheric Disturbance (SID)

During a solar flare, the electron density in the D-region increases by a large factor as a result of the considerable increase in the solar 'hard' x-rays of wavelength less than 10 \AA . This increase is a factor of 100 to 1000 depending on the severity of the flare. Since the increased electron density occurs at altitudes where the electron collision frequency is high, radio waves propagating in the ionosphere are almost completely absorbed, and high-frequency (HF) communications are disrupted over the sun-lit hemisphere. This is sometimes called a radio blackout. A related phenomenon is the PCA event.

Polar Cap Absorption (PCA)

Polar Cap Absorption is the name given to the severe attenuation suffered by a HF radio wave propagating in the ionosphere at a high latitude near the polar caps, in the daytime or at night, soon after a solar flare. It is caused by the high flux of solar protons emitted during a large flare, and deflected to the polar regions by the geomagnetic field.

2 E layer: 90km-140 km:

This is the best understood region of the ionosphere, and the first 'layer' identified in ionospheric research (It was the 'electric' layer - hence the e-layer)

Formation

The E-region is formed primarily by the ionization of O_2 . The solar radiations primarily responsible for the ionization are Lyman- β of wavelength 1025.7 Å ($I_\infty \approx 3.6 \times 10^9$ photons/cm² sec), and the C_{III} line of wavelength 977 Å ($I_\infty \approx 4.4 \times 10^9$ photons/cm² sec). An additional production process is the ionization of N_2 (and O_2) by X-rays of wavelength in the range 10-100 Å. The N_2^+ ions are converted to O_2^+ and NO^+ ions by rapid charge exchange. The net charge production rate is about 4000 electrons/cm³ sec at 105 km for $\chi = 10^\circ$. At high latitudes, particle radiation makes a significant contribution to the ionization at all hours.

The dominant ions in the E-region are O_2^+ and NO^+ , so that the dominant loss process is the dissociative recombination of the electrons with these ions.

Electron Density

The peak electron density around noon, at equinox at the equator (i.e., $\chi = 0$) is $\cong 2 \times 10^5$ el/cm³. The diurnal, seasonal and latitudinal variation is in approximate agreement with the Chapman theory. The height of the peak varies with χ in agreement with the theory, with $h_0 \cong 105$ km and $H \cong 8$ km. Chapman theory shows that the production rate, defined by q_{\max} , is linearly proportional to $\cos \chi$. If the dominant loss process is dissociative recombination, i.e., $q = \alpha N_e^2$, then $N_{e \max}$ should vary as $\cos^{0.5} \chi$. The maximum electron density is found by experiment to vary with χ as $\cos^{0.6} \chi$. The slight difference in the exponent (0.6 versus 0.5) can be accounted for by the height variation of the scale height and of the recombination coefficient. Note that the functional dependence implied by χ can be either time of day or latitude.

Based on a large number of measurements, the solar-cycle variation of the electron density may be expressed by:

$$(N_e)_{\max} \cong a (1 + 0.004 R_Z) \quad (\text{Eqn. 7.12})$$

where R_Z is the sunspot number and $a \cong 1.3 \times 10^5$ el/cm³ at $\chi = 0^\circ$ ('a' varies slightly from month to month)

The E-region persists even during the night, with electron densities in the range 500-10,000 electrons/cm³. The nighttime E-region is thought to be maintained by solar extreme ultraviolet (EUV) radiation, primarily Lyman- α and Lyman- β which has been scattered from the exospheric hydrogen - e.g. the geocoronal glow.

Disturbances (Anomalies)

Within the E-region, local enhancements in electron density are frequently observed. These are known as sporadic-E, or E_s . Ground-based observations (global network of ionosondes) show that E_s is more prevalent in summer than in winter, at mid-latitudes. At the geomagnetic equator (actually the magnetic dip equator), E_s is observed mainly in the daytime, throughout the year. Rocket-borne experiments have shown that, at mid-latitudes, E_s is a thin layer, of thickness in the order of a few hundred meters, with N_e greater than $(N_e)_{\max}$ of the E-layer. The processes involved in the production of E_s are rather complicated.

3 F-region: 140km-1000km

This is the region that is primarily responsible for the reflection of radio waves in high-frequency communication, broadcasting, and OTHR (over-the-horizon radar) - hence the most important of the ionospheric regions.

Formation

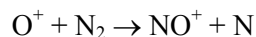
The primary production process in the F-region is the ionization of atomic oxygen, O, by solar radiation of $\lambda < 911 \text{ \AA}$. The spectral bands responsible for the ionization are the Lyman continuum: $800 - 910 \text{ \AA}$ ($I_\infty \cong 1.0 \times 10^{10} \text{ ph/cm}^2 \text{ sec}$); the wavelength range $200-350 \text{ \AA}$, including the strong He II line at 304 \AA ($I_\infty \cong 1.5 \times 10^{10} \text{ ph/cm}^2 \text{ sec}$); and the wavelength range $500-700 \text{ \AA}$. A secondary production process is the ionization of molecular nitrogen and oxygen by solar radiation of $\lambda < 796 \text{ \AA}$. Peak electron production occurs in the height range of $\cong 160-180 \text{ km}$, but the peak N_e occurs at a greater height (see below).

The primary positive ions produced by the above radiations are O^+ , N_2^+ and O_2^+ . Various chemical process then convert these ions into different ions, so that the observed dominant ions are:

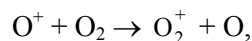
O^+ , NO^+ and O_2^+ in the height range 140-200 km (F_1), and
 O^+ in the height range 200-400 km (F_2).

The F-region divides into two layers, called F_1 and F_2 , particularly in the summer in the daytime. The F_1 -layer, forms a ledge in the electron density profile at the bottom of the F_2 -layer.

The main loss process in the F_1 -layer ($\cong 140-200 \text{ km}$) is the dissociative recombination of the electrons with the molecular positive ions. In the F_2 -layer ($\cong 200 \text{ km}$ to $\cong 400 \text{ km}$) the main chemical loss process is a two-stage process in which ion-molecule reactions first convert O^+ to the molecular ions NO^+ and O_2^+ by the reactions



and



and the electrons then recombine dissociatively with these ions. This loss process obeys a linear law, i.e., $L = \beta(h) N_e$, with the loss coefficient $\beta(h)$ decreasing with height faster than the production $q(h)$ decreases with height, so that the electron density increases with height above the level of peak production (as noted earlier). The F_2 -peak is formed by the combined action of this linear loss process and the transport of electrons by plasma diffusion. In fact, the transport of electrons by various processes plays an important role in the morphology of the F_2 -layer.

Above 500 km, significant amounts of the hydrogen ion H^+ are observed and, at heights greater than about 1200 km, H^+ is the dominant ion, especially at night.

Electron Density

The maximum electron density of the F₁-layer is about 3×10^5 el/cm³, around noon. The layer disappears at night. The peak of the F₂-layer is about an order of magnitude higher in density. The diurnal, seasonal and latitudinal variation of the electron density and height of the F₂ - layer are NOT in accordance with the χ -variation. For instance, $(N_e)_{\max}$ is:

(a) greater in winter (2.5×10^6 at $R_Z = 200$) than in summer (7×10^5 at $R_Z = 200$). This is called the 'winter anomaly',

(b) greater on either side of the equator than at the equator - the Appleton (equatorial) anomaly, and

(c) greater before noon, or afternoon, than at noon in the summer.

Points a and c are illustrated in Figure 7.6, which shows how winter densities are higher than summer, and shows the increase in electron density with increasing solar activity.

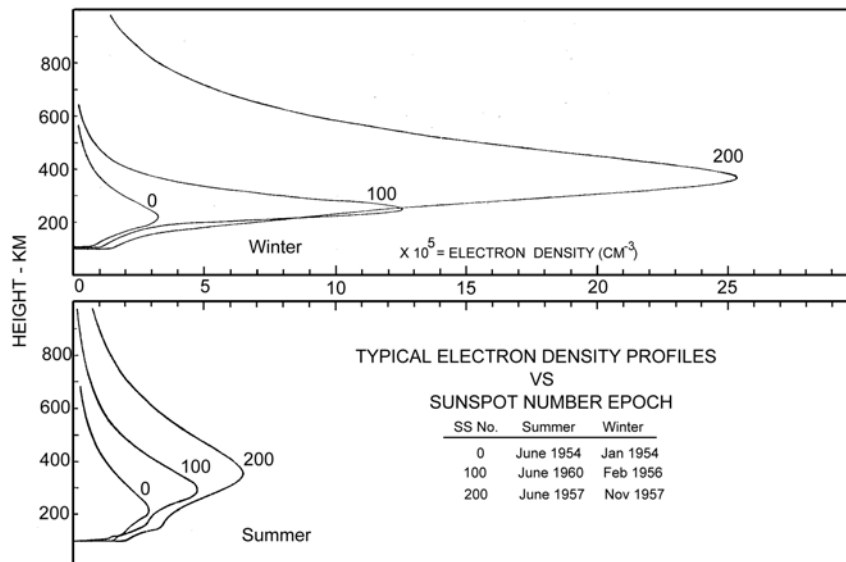


Figure 7.6 Typical electron density profiles at three (Zurich) sunspot epochs for winter and summer noon conditions at Washington DC (Belvoir). Magnetically quiet days were chosen for the above profiles. J. W. Wright, Dependence of the Ionospheric F Region on the Solar Cycle, *Nature*, page 461, May 5, 1962.

Disturbances (Anomalies)

The departures from a simple solar control of the electron density in the F₂-region are referred to as anomalies. These anomalies are the result of several factors:

- a) marked seasonal and solar-cycle variation of the temperature and consequently the scale height H,
- b) seasonal variation of the O/O₂ and O/N₂ ratios,
- c) transport of electrons by diffusion, winds (global atmospheric circulation), electromagnetic forces ('fountain effect') etc.
- d) transport of electrons between geomagnetic conjugate points.

The F₂-region is also disturbed by solar flares. About 24-36 hrs after a solar flare, particle radiation from the sun causes $(N_e)_{\max}$ in the F₂-layer to increase for a few hours at high latitudes. This is followed by a decrease which lasts for several hours. Such ionospheric disturbances could seriously affect high-frequency communication and broadcasting, as well as OTHR.

D Ionosphere morphology - Latitude Structure

The ionospheric structure described above is descriptive of the mid-latitude ionosphere. Near the magnetic equator, and at high latitudes, there are substantial deviations from the simple forms described by Chapman theory. These are largely due to the influence of the magnetosphere on the earth's upper atmosphere.

The most prominent feature of the ionosphere is the aurora, illustrated at the beginning of the chapter in Figure 7-1. The aurora occur at high latitudes - typically near 60-65° magnetic latitude. The bright auroral display is most obvious to the human eye (and satellites such as DMSP) near local midnight. In fact, the aurora occurs in an oval, or ring shaped distribution centered on the magnetic poles, and extends to all longitudes. Figure 7-7 shows an example of the Dynamics Explorer 1 imagery which shows the auroral oval extending around the northern magnetic pole. When the DE-1 imagery emerged in late 1981, it confirmed concepts such as the auroral oval with remarkable scenes such as the one shown here, and demonstrated for the first time how the aurora extended into the dayside polar region.

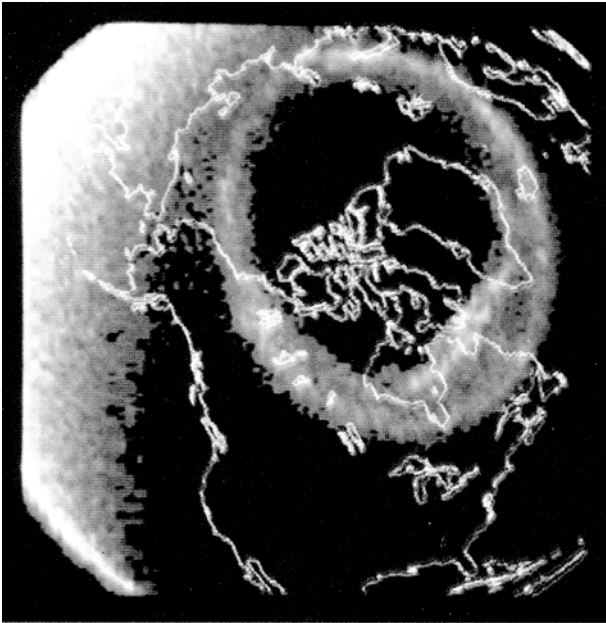


Figure 7.7 Auroral imagery from DE-1. This image was taken on 8 November 1981 by the Dynamics Explorer 1 Spin-Scan Auroral Imager. The filter used here passed the O_I wavelengths of 130.4 and 135.6 nm (far UV). A coastline map was superimposed on the image, which was taken at an altitude of several earth-radii above the northern polar cap. The illuminated hemisphere is to the left - the polar region is largely in darkness. These images were the first to show an uninterrupted look at the entire auroral oval.

Figure is from Plate 1 of, *Images of the Earth's Aurora and Geocorona from the Dynamics Explorer Mission*, L. A. Frank, J. D. Craven, and R. L. Rairden, *Advances in Space Research*, vol. 5, No. 4, pp. 53-68, 1985.

These auroral displays are produced by electrons from the plasma sheet, precipitating along magnetic field lines which extend in the night sector to $L = 5$ to 10. Some of the most intense displays map to geosynchronous orbit. The auroral electrons typically have energies of 1-10 keV, and often have a mono-energetic spectrum which in recent years has been interpreted as due to acceleration by electric fields pointed along the auroral magnetic field lines. (McIlwain first inferred this from sounding rocket measurements in 1960.) Figure 7-8 shows the topology described here.

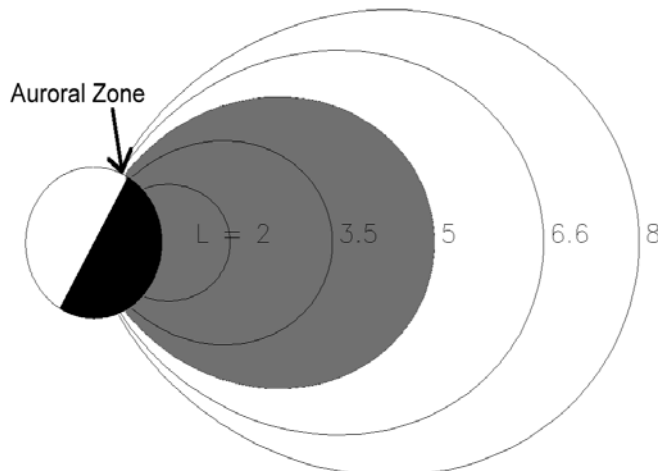


Figure 7.8. Diagram of auroral field lines. The sun is to the left. The gray shaded region is meant to illustrate the plasmasphere. Outside this region ($L > 5$ typically), is the plasma sheet. Plasma sheet electrons with energies of 1-20 keV move along the magnetic field lines. Those which fall into the (magnetic) loss cone precipitate into the atmosphere, producing a diffuse auroral glow. At the high latitude edge of this region, more intense aurora can be seen (visible to the human eye). This 'discrete' aurora is thought to result when plasma sheet electrons are accelerated to higher energies by electric fields aligned along the magnetic fields within a few thousand kilometers of the earth's surface. Magnetic fields poleward of this region map to the tail lobes, and perhaps out into the solar wind.

There is also a substantial structure at lower latitudes. Figure 7.9 shows unique imagery from early in the space age - the UV camera carried to the moon on Apollo provided a unique look at the earth's atmosphere. Thin bands of light are seen to extend from the daylight hemisphere into the dark hemisphere. These bands reflect the equatorial structure associated with the equatorial fountain - aka the Appleton Anomaly.

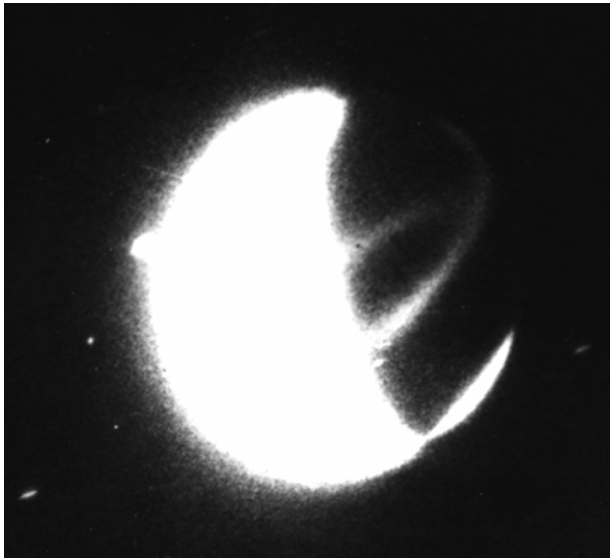


Figure 7.9. UV image of Earth, April 21, 1972. This image was recorded with the far UV camera carried on the Apollo 16 spacecraft. The image was acquired with a pass-band of 1250-1600 Å. The exposure time was 10 minutes. The bright regions are from atomic oxygen and molecular nitrogen emissions in the upper atmosphere. (Note the hydrogen geocorona is not visible in this image.) The sun is to the lower left, producing the bright airglow in the sunlit hemisphere. The southern auroral zone is the bright cap to the lower right; there is a less obvious brightening in the partially sunlit northern hemisphere auroral zone. The equatorial airglow bands extend into the night, curving from $\sim \pm 20^\circ$ latitude to near the equator at the right side.

George R. Carruthers and Thornton Page, Apollo 16 Far-Ultraviolet Camera/Spectrograph; Earth Observations, *Science*, 1 September 1972, volume 177, pp 788-791.

This remarkable image reflects an important latitudinal structure in electron density. Figure 7.10 shows this structure, and the substantial change in electron density observed near the equator. This structure is due to the solar driven upwelling of the neutral atmosphere at the equator, which in turn forces an upward flow of the plasma there. This has been termed the equatorial fountain. The plasma drifts back downward 5-10 degrees away from the equator. These density variations, and the turbulence associated with the equatorial fountain, make the low latitudes a region in which it is difficult to utilize equipment such as OTHR.

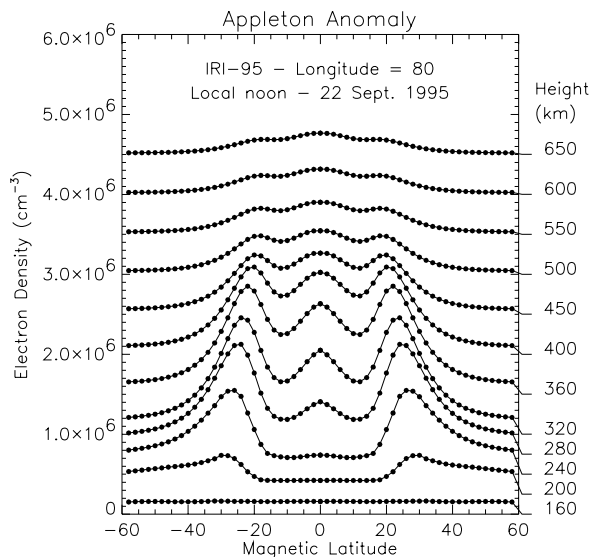


Figure 7.10 Equatorial density profiles. The profiles shown here were obtained from the IRI-95 model, run for 22 September 1995. The resulting latitude profiles were plotted beginning at the bottom with a profile at 160 km altitude. Subsequent plots were offset upward by 0.25×10^6 (up to 320 km) and then by 0.5×10^6 . The zero points are indicated by the short horizontal lines leading from the plot to the altitude labels

E Radiowave propagation in the Ionosphere

When an electromagnetic wave is incident on the ionosphere the oscillating electric field of the wave causes the electrons to oscillate with respect to the much heavier and therefore relatively stationary positive ions. This process results in the refraction or reflection of radiowaves (depending on their frequency) from the ionosphere as we shall see in more detail below. In general the influence of the ionosphere on wave propagation decreases with increasing frequency and waves of frequency higher than the so called "plasma" frequency will pass through the ionosphere and reach a satellite. Because of its reflective properties at sufficiently low radio frequencies the ionosphere behaves like a "mirror in the sky" and is used for short wave communications, OTHR (over the horizon radar) and radio navigational systems.

High frequency (3 - 30 MHz) radio waves are reflected from the dayside E region or the F region (200-300 km altitude) while very low frequencies (less than 30 kHz) are reflected from the D region (70-100 km altitude). Radio waves are also partially absorbed by the ionosphere as a result of electron collisions with air molecules. This is particularly true in the D region where the density of neutral air molecules is largest. The increased D region ionization caused by an intense solar flare can result in the complete absorption of high frequency waves and disruption of short wave communication, particularly over the sunlit hemisphere of the earth.

To understand how a radiowave is reflected from the ionosphere we need to develop a few concepts:

(1) The refractive index of a medium n is defined as: $n = \frac{c}{v_p}$

where :

c = the velocity of an electromagnetic wave in a vacuum

$= 3 \times 10^8$ m/s, and

v_p = the phase velocity of the wave in the medium

In most substances such as glass or water $n > 1$ and n varies with frequency. It is greater for violet light (high f) and smaller for red light (lower f).

In a plasma medium the index of refraction (neglecting the geomagnetic field) is:

$$n = \sqrt{1 - \frac{Ne^2}{\epsilon_0 m \omega^2}} = \sqrt{1 - \frac{80.6 N}{f^2}} \quad (\text{Eqn. 7.13})$$

where N = electron density in (electrons/ m^3) and f is the frequency of the wave in Hz. (In this section of the text, we must deviate from the previous practice of using the lower case "n" for electron density, since "n" is used for the index of refraction.)

This formula exhibits a characteristic frequency which determines the boundary between a real and an imaginary index of refraction. This characteristic frequency, known as the plasma frequency, is defined by the equation

$$f_p = \sqrt{80.6 N} \quad \text{or} \quad \omega_p = \sqrt{\frac{Ne^2}{\epsilon_0 m}} \quad (\text{Eqn. 7.14}).$$

This represents the natural frequency of oscillation of the electrons around the stationary ions. If, in contrast with the behavior of light in dielectric materials such as glass. If , the index of refraction is imaginary. This means that in the plasma medium, electromagnetic waves cannot propagate below a critical frequency for which the medium is conducting.

(2) Snell's Law tells us how a wave is refracted (bent) at the interface between two media.

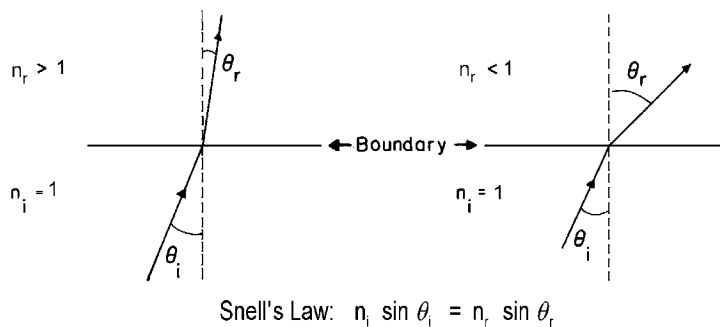


Figure 7.8 Snell's Law - Single Boundary

Consider now a layered medium in which each layer has a lower index than the one below it as shown.

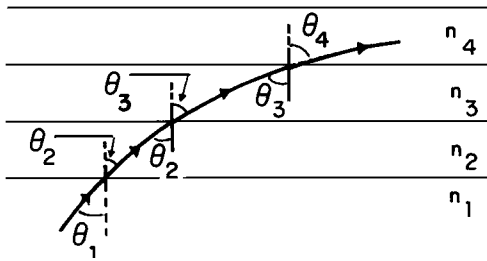


Figure 7.9 Snell's Law - Continuously varying index of refraction

Applying Snell's Law at each boundary we obtain

$$\frac{n_2}{n_1} = \frac{\sin \theta_1}{\sin \theta_2} \quad \frac{n_3}{n_2} = \frac{\sin \theta_2}{\sin \theta_3} \quad \frac{n_4}{n_3} = \frac{\sin \theta_3}{\sin \theta_4} \text{ etc.}$$

This can also be written in the form

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 = n_3 \sin \theta_3 = \dots \text{or in general}$$

$$n \sin \theta = \text{constant} \quad (\text{Eqn. 7.15})$$

Let us now apply this to the ionosphere as shown in figure 7.10

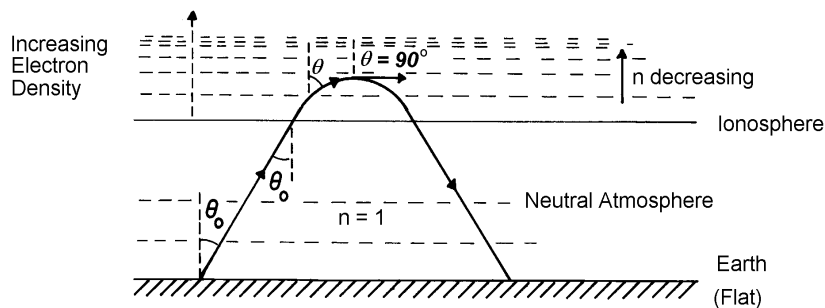


Figure 7.10 Snell's Law in the ionosphere

Eqn. 7.15 must hold for all points of the ray path. Before the wave enters the ionosphere $\theta = \theta_0$ and $n = 1$. Thus the angle of incidence at the base of the ionosphere is also θ_0 as can be seen from the sketch.

As we enter the ionosphere the electron density increases and the index of refraction begins to decrease below 1. To satisfy Snell's Law the angle θ must increase and the ray bend further and further away from the vertical. In order to have reflection take place θ must go to 90° at the top of the trajectory. We can calculate what the electron density must be for this to take place

$$n_m \sin 90^\circ = (1) \sin \theta_0 \quad \text{or} \quad n_m = \sin \theta_0 \quad (\text{Eqn. 7.16})$$

But we had above in eqn.s 7.13 and 7.14:

$$n = \sqrt{1 - \frac{N e^2}{\epsilon_0 m \omega^2}} = \sqrt{1 - \frac{80.6 N}{f^2}} = \sqrt{1 - \frac{\omega_p^2}{\omega^2}} = \sqrt{1 - \frac{f_p^2}{f^2}} \quad (\text{Eqn. 7.17})$$

Thus the minimum electron density N_m needed to reflect this ray which left the ground at an angle θ_0 is given by

$$\sin \theta_0 = \sqrt{1 - \frac{80.6 N}{f^2}} = \sqrt{1 - \frac{f_p^2}{f^2}} \quad (\text{Eqn. 7.18})$$

where f = frequency of wave (Hz)

Solving for N_m we obtain (using $1 - \sin^2 \theta_0 = \cos^2 \theta_0$)

$$N_m = \frac{f^2}{80.6} \cos^2 \theta_0 \quad (\text{electrons/m}^3) \quad (\text{Eqn. 7.19})$$

If we now consider the special case of vertical incidence we have that $\theta_0 = 0$ which gives for the minimum required electron density N_0

$$\sqrt{1 - \frac{80.6 N_0}{f^2}} = \sqrt{1 - \frac{f_p^2}{f^2}} = 0 \quad (\text{Eqn. 7.20})$$

Thus we see that for vertical incidence $n = 0$ when the incident wave frequency equals the plasma (or critical) frequency.

To summarize: For vertical incidence reflection will occur at a level in the ionosphere where

- | | |
|--------------------|--|
| (a) $n = 0$ | Index of refraction vanishes |
| (b) $f = f_p$ | radio frequency = plasma frequency |
| (c) $v_p = \infty$ | Phase velocity becomes infinite |
| (d) $v_g = 0$ | Group velocity = velocity of energy transport becomes 0. |

If we now compare the minimum electron density required for reflection for the vertical and oblique cases we see that

$$\text{Oblique:} \quad N_m = \frac{f^2}{80.6} \cos^2 \theta_0 \quad (\text{Eqn. 7.21 a})$$

$$\text{Vertical:} \quad N_m = \frac{f^2}{80.6} \quad (\text{Eqn. 7.22 b})$$

A smaller electron density is required for oblique incidence ($\cos \theta$ is always ≤ 1)

If we look at the same problem from the point of view of the frequencies which will be reflected for given electron density at a certain height we find that the frequency f_θ of the oblique wave is related to the reflected frequency f_p at vertical incidence by

$$f_\theta = \frac{f_p}{\cos \theta} = f_p \sec \theta \quad (\text{Eqn. 7.23})$$

which shows that in general $f_\theta > f_p$.

If N_{\max} represents the maximum (peak) electron density in an ionospheric layer then the highest frequency that will be reflected at vertical incidence is called the "critical" frequency of the layer and is given by

$$f_o = \sqrt{80.6 N_{\max}} \approx 9\sqrt{N_{\max}} \text{ with } N_{\max} (\text{electrons/m}^3) \quad (\text{Eqn. 7.24})$$

and following our earlier discussion we see that for oblique incidence the critical frequency will be

$$(f_\theta)_c = f_o \sec \theta \quad (\text{flat earth}) \quad (\text{Eqn. 7.25})$$

For large values of θ (approaching 90°) this leads to a substantial increase in the available frequencies for communications. Note that this formula does not take into account the earth's curvature, and further modifications are necessary for large θ .

The applicability of this result is illustrated in figure 7.11, an ionospheric sounding utilizing radar at frequencies of 1-10 MHz. The radar is swept in frequency (horizontal axis), and the range-gated returns (vertical axis) give the altitude at which the signal is reflected.

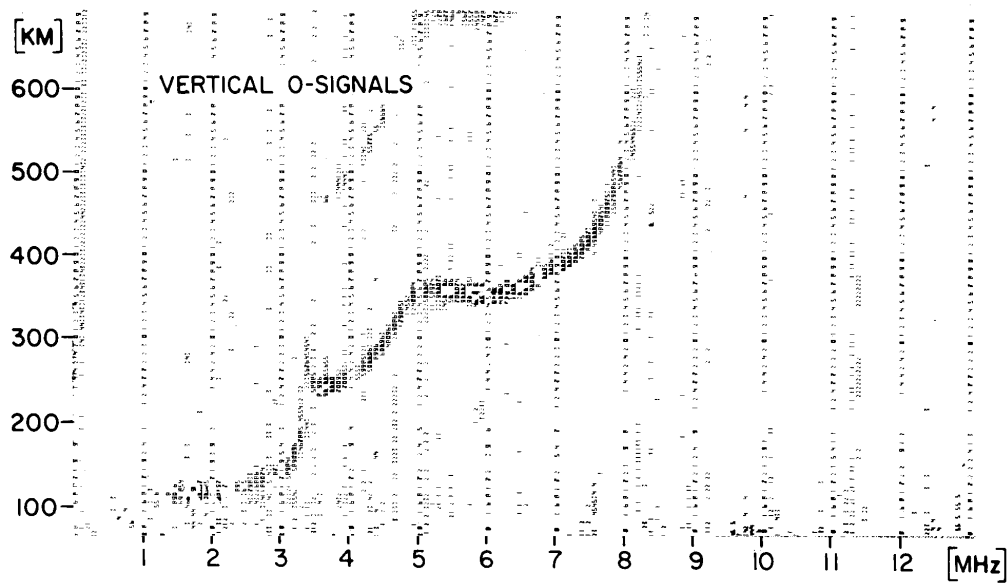


Figure 7.11 Data from a digital ionosonde at the AFGL Goose Bay Ionospheric observatory. These data were acquired on 16 June 1980 at 1720 AST. (local noon?) There is a standard set of 128 range bins at each frequency. Height resolution is 5 km, each frequency step is 100 kHz. Handbook of Geophysics, page 10-6. See also Bibl, K. and B. W. Reinisch, The universal digital ionosondes, Radio Science, vol. 13, pages 519-530, 1970.

F References

Friedman, Sun and Earth, Scientific American Press, New York, 1986

Kelley, Michael C, The Earth's Ionosphere, Academic Press, San Diego, 1989.

McIlwain, C. E., Direct measurement of particles producing visible auroras, Journal of Geophysical Research, 65, page 2727, 1960.

Ratcliffe, J. A., An Introduction to the ionosphere and magnetosphere, Cambridge University Press, 1972.

Handbook of Geophysics and the Space Environment, edited by Adolf S. Jursa, Air Force Geophysics Laboratory, 1985; Chapter 10, The Ionosphere, Section 10.1 Measuring Techniques, S. Basu, J. Buchau, F. J. Rich, and E. J. Weber

G Problems

1. Determine the production curve resulting from the photoionization of an atmospheric gas by monochromatic solar radiation incident vertically on the atmosphere. The rate of electron production at the altitude h is given by:

$$q(h) = I(h) \sigma n(h) \text{ electrons/m}^3 \text{ s}$$

where

$I(h)$ = intensity of the solar radiation (in photons/m²s) at the altitude h

σ = ionization cross-section (in m²) of the gas molecules for the particular solar radiation. and

$n(h)$ = number density (in m⁻³) of the gas molecules at the altitude (h)

(a) For n , use the barometric relation with a scale height of 30 km and $n = 5 \times 10^{16}$ molecules/m⁻³ at 150 km altitude

(b) For the solar radiation, use $I(h) = I_0 e^{-(1000 - h)/h_0}$, h_0 is the scale height for the decrease in solar intensity with altitude. Take $I_0 = 1.0 \times 10^{14}$, h_0 (km) = $10^{17}/n$. note that the last formula

should be read: $h_0 (km) = \frac{10^{17}}{\text{density}}$

(c) For the ionization cross section, take $\sigma = 2 \times 10^{-21}$

Plot the neutral density, $n(h)$, solar intensity, $I(h)$, and the production rate, $q(h)$ as a function of altitude for $h = 250$ to 800 km altitude.

2. Figure 7.12 shows an electron density profile. Estimate the scale height for plasma above 700 km altitude (Go from 700-1500 km). Use this scale height to estimate the temperature. Note that the answer depends upon which mass you use in the scale height formula $\left(H = \frac{kT}{mg}\right)$. Try the

electron mass, proton mass, and the mass of atomic oxygen (16 AMU). Which is most reasonable?

3. From Figure 7.12 estimate the electron density at 200 km and 400 km. What are the radio frequencies that will be reflected at these altitudes assuming vertically traveling waves?

4. Compare the neutral gas density and plasma (electron) density as a function of altitude. Plot the total neutral density (number/volume) as a function of altitude (e.g. figure 6.5b). On the same curve, plot the electron density.

5. Using the data found in Figure 7.11, estimate the electron density profile. ($\theta = 0$)

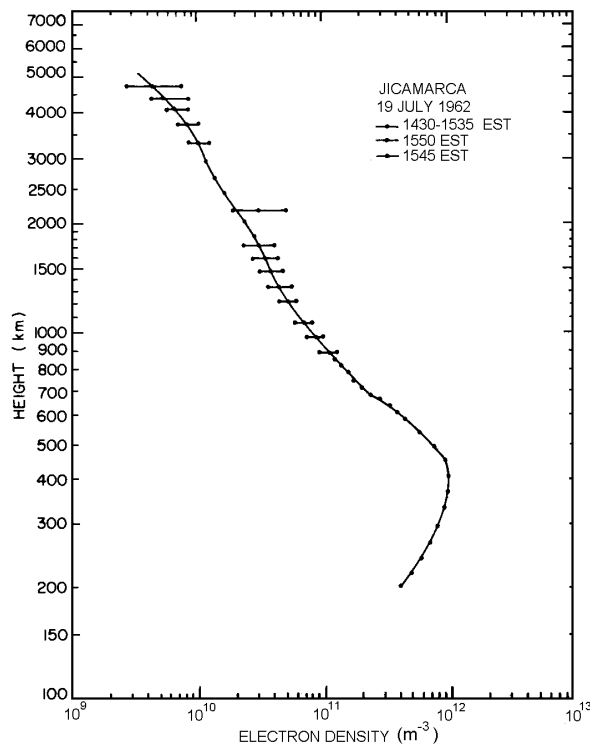


Figure 7.12 An electron density profile obtained at Jicamarca that extends to almost one earth radius. [Bowles, K. L. and Staff of Jicamarca Radio Observatory Institute, Profiles of Electron Density Over the Magnetic Equator Obtained Using the Incoherent Scatter Technique, *Tech Report Note 169*, National Bureau of Standards, Boulder, CO, 1963]. [Found in Handbook of Geophysics, page 10-16]

6. The atmosphere of a hypothetical planet is composed of nitrogen, N_2 , and oxygen, O_2 . Above 150 km, the atmosphere is isothermal at a temperature of 1500 K, and its composition varies with altitude as a result of diffusive separation.

The number densities of N_2 and O_2 are $5 \times 10^8 \text{ cm}^{-3}$ and $5 \times 10^7 \text{ cm}^{-3}$ respectively at 200 km; and their **absorption** cross-sections for solar radiation of wavelength 75 nanometers are $3.1 \times 10^{-17} \text{ cm}^{-2}$ and $2.9 \times 10^{-17} \text{ cm}^{-2}$ respectively.

The **ionization** cross-section of N_2 for the same radiation is $2.3 \times 10^{-17} \text{ cm}^{-2}$.

The acceleration due to gravity is 9.81 m/s^2 , and may be assumed to be independent of height.

If the solar irradiance at the top of the atmosphere is $2.5 \times 10^9 \text{ photons/cm}^2 \text{ s}$ at 75 nm, what would be the irradiance at 200 km when the radiation is vertically incident on the atmosphere?

Calculate the ionization rate of N_2 at 200 km due to the 75 nm radiation.

Assuming that the only loss process for N_2^+ is charge transfer to O_2 , and that this reaction rate is $1.5 \times 10^{-11} \text{ cm}^3 \text{ s}^{-1}$ calculate the PCE density of N_2^+

Take $\gamma = 1$ (one electron per incident photon).

Blank page

Chapter 8 Environmental Effects on Space Systems

A Spacecraft Charging

1 Introduction

Spacecraft charging results from the interaction of satellites with charged particles and energetic photons in the space environment. Two classes of phenomena are of interest:

- (a) Surface charging
- (b) Internal charging

These processes lead to a variety of operational anomalies as listed.

- (a) Degradation of surfaces including sensor surfaces
- (b) Interference in electrical circuits due to arcing
- (c) Internal charge accumulations in electronic components

Some of these effects can have serious consequences including loss of the spacecraft or mission objectives.

2 Charging Mechanisms

While there are several mechanisms which contribute to spacecraft charging, the three most important ones are

- (a) Charge flow from the ambient plasma
- (b) Photoelectron emission from the spacecraft
- (c) Secondary emission due to plasma bombardment

These currents produce a current flow to the satellite, and generally result in a non-zero charge on the satellite body (and surfaces), and hence non-zero potentials. These potentials can be *quite* large. The general scenario is presented in figure 8.1, where the different current sources are identified.

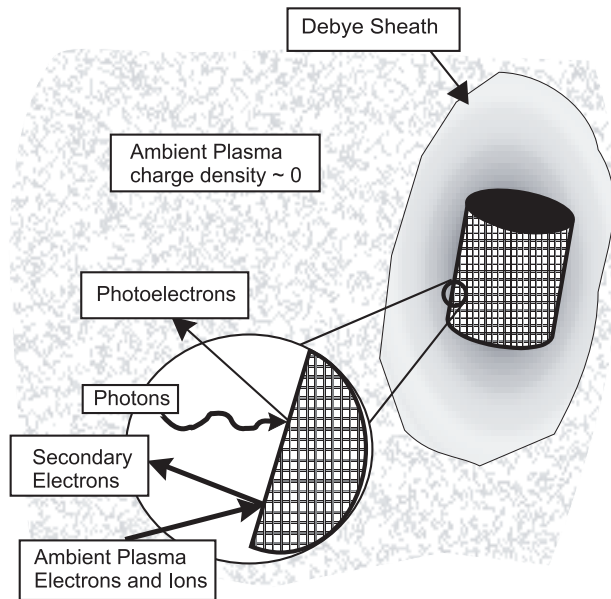


Figure 8.1. Spacecraft Charging Processes. Note that the charge distribution near the satellite can be non-neutral, and in the case of actively biased systems (such as near solar arrays), extremely non-neutral.

Plasma Induced Charging

In a plasma, charge neutrality is maintained by strong, long range electric forces. This generally requires that on average, the electron and (positive) ion densities are equal (quasi-neutrality). The average kinetic energies (temperatures) are often similar. In general, their velocities are quite different, however. This can be seen if we go ahead and assume that both electrons and ions (usually protons) are at the same temperature.

The thermal velocity is given by:

$$\langle E \rangle = \frac{3}{2} kT = \frac{1}{2} m_e v_e^2 = \frac{1}{2} m_i v_i^2. \quad (\text{Eqn. 8.1})$$

Assuming that most ions are protons, we see that the ratio of thermal velocities is fairly high:

$$\frac{v_e}{v_i} = \sqrt{\frac{m_i}{m_e}} = \sqrt{\frac{m_p}{m_e}} = 43 \quad (\text{Eqn. 8.2})$$

Hence the electron velocity is much higher than the ion velocity. In the simplest case (no effects due to satellite motion or plasma flow), the current density (current/area) incident on a satellite surface is given by:

$$j = qnv \quad (\text{Eqn. 8.3})$$

where:

q = charge on particle (Coulombs)

n = particle density (number/m³)

v = (thermal) velocity (m/s)

Since the charge and density are the same, we see that the current density is much higher for the electrons than the ions (43 times higher, for a proton plasma). (Study question - what is the ratio for an oxygen (O+) environment ?)

Example

Typical values of $kT_e = kT_i = 8000\text{eV}$ at Geosynchronous orbit. Typical densities are about 10^6 particles/ m^3 and $q = \pm 1.6 \times 10^{-19}\text{C}$, yielding for the velocities

$$v_e = \sqrt{\frac{kTe}{m_e}} = \sqrt{\frac{8000 \times 1.6 \times 10^{-19}}{9.1 \times 10^{-31}}} = 3.75 \times 10^7 \text{ m/sec} \quad (\text{Eqn. 8.4a})$$

$$v_p = \sqrt{\frac{kTe}{m_p}} = \sqrt{\frac{8000 \times 1.6 \times 10^{-14}}{1.6 \times 10^{-27}}} = 8.67 \times 10^5 \text{ m/sec} \quad (\text{Eqn. 8.4b})$$

and the resulting current densities are:

$$j_e = -n_e e v_e = -(10^6)(1.6 \times 10^{-19})(3.75 \times 10^7) = -6 \times 10^{-6} \left(\frac{\text{A}}{\text{m}^2} \right) \quad (\text{Eqn. 8.5a})$$

$$j_p = +n_p e v_p = +(10^6)(1.6 \times 10^{-19})(8.67 \times 10^5) = +1.3 \times 10^{-7} \left(\frac{\text{A}}{\text{m}^2} \right) \quad (\text{Eqn. 8.5b})$$

for the electrons and protons incident on the spacecraft

Typical cross-sectional areas are 5 m^2 , so the net current is

$$\begin{aligned} I_{\text{net}} &= (j_{\text{net}})(\text{Area}) = (j_e + j_p)(\text{Area}) \\ &= (-6 \times 10^{-6} + 1.3 \times 10^{-7})(5) = -2.9 \times 10^{-5} (\text{Amp}) \end{aligned} \quad (\text{Eqn. 8.6})$$

Because of this negative charge accumulation, a net potential develops on the satellite. Using our example, in one millisecond, a charge $\Delta q = (I_{\text{net}})(t) = -2.9 \times 10^{-8} \text{C}$ develops. If the capacitance is 100 pF (a typical value), a potential of $V = q/C = -0.29 \text{ kV}$ is developed. (The capacitance can be estimated by assuming the satellite is a sphere - see Halliday and Resnick - $C(\text{picoFarads}) \sim \text{radius}(\text{cm})$)

This potential reduces the incident electron flux until it is balanced by the ion flux. The mathematics of this process is fairly simple, but one must do a little kinetic theory to obtain the effect of the potential on the currents. For the repelled specie ($q\Phi > 0$):

$$I = I_0 e^{q\Phi/kT} \quad (\text{Eqn. 8.7a})$$

and for the attracted specie ($q\Phi < 0$):

$$I = I_0 \left(1 - \frac{q\Phi}{kT} \right) \quad (\text{Eqn. 8.7b})$$

where $I_0 = qnvA$.

Equilibrium occurs when there is no net current - then the potential stops changing. This point can be determined by setting the two terms for current above equal to each other. Roughly, this tends to occur when the potential energy of the electrons is roughly equal to their average kinetic energy.

$$-e\Phi_s \cong kT \Rightarrow \Phi_s = -\frac{kT}{e} \quad (\text{Eqn. 8.8})$$

where Φ_s = spacecraft potential (Volts). The electron flux is then substantially reduced, and the ion current is increased slightly. This relation holds reasonably true in the absence of other current sources, which as a practical matter tends to mean LEO, in eclipse.

Otherwise, there are several additional current sources which are important. The largest of these is photoemission.

Photoelectric Emission

When photons of sufficient energy strike a material surface electrons are ejected from the surface leaving the target positively charged. This process can be represented as shown

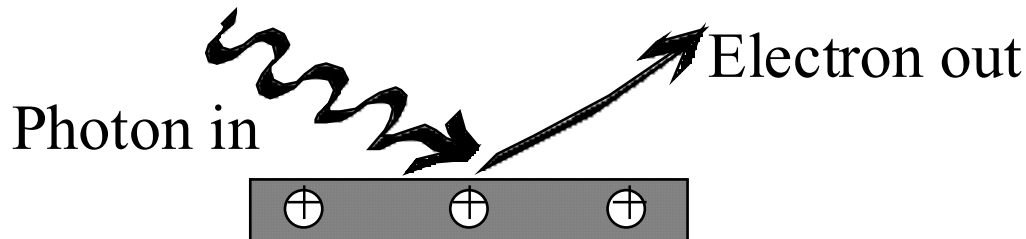


Figure 8.2 Photoemission

The energy balance in this process can be written

$$hf = \phi + KE_e \quad (\text{Eqn. 8.9})$$

where hf = Energy of incident photon

ϕ = Work function of the material

KE_e = Kinetic energy of ejected electron

As can be seen in Table 1 the work functions of most spacecraft materials are about 4-5 eV which means that only photons in the UV and X-Rays region ($\lambda \leq 300\text{nm}$) can generate photoelectrons. Nevertheless because of the intense fluxes of UV and soft X-Rays (particularly the Lyman-alpha) in the solar spectrum this is an important mechanisms for positive charging of spacecraft. Since this mechanism is operative only when the spacecraft is in sunlight it results in a cycling of the spacecraft charge for orbiting satellites that pass in and out of eclipse. (Compare the values in Table 8.1 to the current densities calculated in equation 8.5.)

Table 8.1

Material	Work Function (eV)	Photoelectron Saturation Flux ($10^{14}/\text{sec-m}^2$)	Saturation Current Density ($\mu\text{A/m}^2$)
Aluminum Oxide	3.9	260	42
Indium Oxide	4.8	190	30
Gold	4.8	180	29
Stainless Steel	4.4	120	20
Aquadag	4.6	110	18
LiF on Au	4.4	90	15
Vitreous Carbon	4.8	80	13
Graphite	4.7	25	4

Secondary Emission

Finally, there are substantial currents generated by the incident particles. Incident electrons with energies of a few hundred electron volts (100-500 eV) will produce secondary electrons. These secondary electrons escape the surface with energies of a few eV, and their energy spectrum resembles that of the photoelectrons. For some materials, the secondary electron yield can exceed one. This can lead to the anomalous result of having a net positive current produced by the incident electron flux. As the plasma temperature increases, this net secondary yield decreases, typically at temperatures exceeding 10 KeV. (That is, one needs a substantial number of electrons with energies over 10 KeV in order for a net negative charge to flow to the satellite, even in eclipse.)

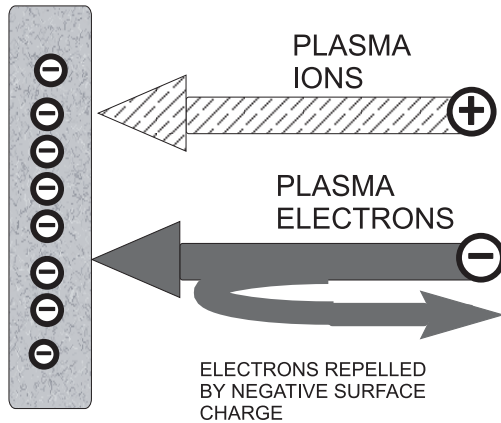
3 Example of Spacecraft Charging – Geosynchronous Orbit

For geosynchronous satellites, spacecraft charging divides itself into two regimes - sunlight and eclipse (shadow). In hot plasma environments, a conducting satellite will experience positive potentials in sunlight, negative potentials in eclipse.

The above discussion assumes that the satellite surface is conducting. This is not typically true, since the materials covering a satellite usually include glass (SiO_2 cover cells on the solar arrays) and kapton (thermal blanket). This is important because shadowed, insulating surfaces can develop large negative potentials (hundreds to thousands of volts), even though the illuminated surfaces and satellite ground have small positive potentials. This “differential charging” can lead to arcing on the satellite surface. These arcs lead to electrical pulses which can couple into the satellite command and control circuitry, leading to anomalous satellite behavior

CHARGING CONDITIONS IN GEOSYNCHRONOUS ORBIT

a) ECLIPSE - NEGATIVE CHARGING



b) SUNLIGHT - POSITIVE CHARGING

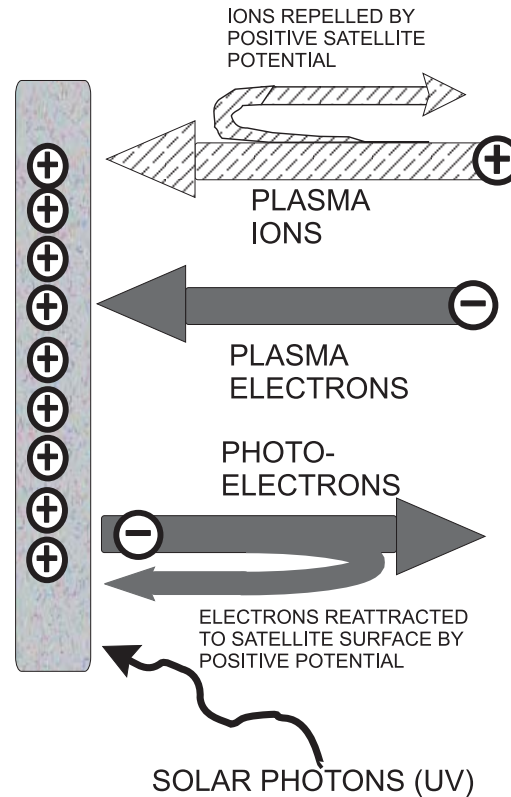


Figure 8.3. Qualitative illustration of the charging of a surface by a plasma. The width of the arrows is proportional to the flux of each particle species; the equilibrium potential is reached when the sum of the currents collected and emitted by a surface element is zero. (a) Surface in shadow: the current balance requires equality between the flow of the plasma ions and that of the electrons impinging on the surface. (b) Surface in sunlight: equilibrium is achieved when the flow of escaping photoelectrons is equal to the difference between the incoming flows of plasma electrons and ions.

As a special illustration of the magnitudes of charging which can be experienced at geosynchronous orbit, the record charging event from Applied Technology Satellite 6 is shown in Figures 8.4 and 8.5. Figure 8.4 is a spectrogram for a two hour interval in 1975, covering an eclipse. This figure is similar in spirit to that shown in chapter 5, but the energy range is substantially larger, extending upwards to 80 keV. The eclipse begins just prior to 2100, and ends just before 2200. There is an injection of hot plasma about 10 minutes into the eclipse, and the satellite potential reaches -19 kV. Other satellites have not been observed to charge to such levels - we were just lucky.

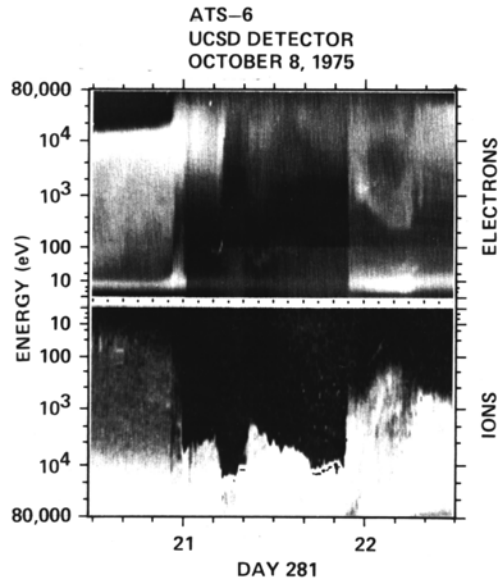


Figure 8.5. Ion and electron spectrogram for the record charging event. Note that by definition, the event occurs at local midnight, typically a region of relatively hot plasmas. The electron energies reach (and apparently exceed) 80 keV.

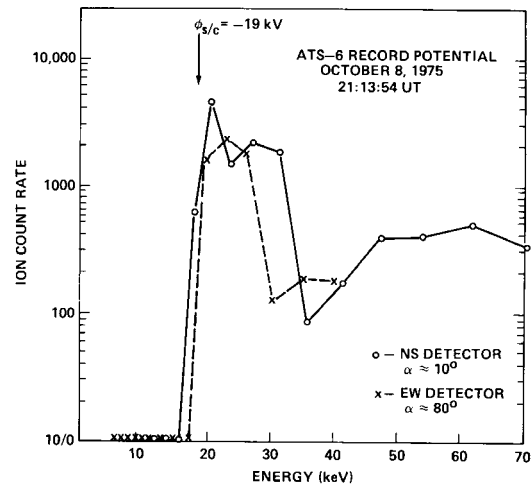


Figure 8.5. Ion spectra for the record charging event.

4 Design Considerations

There are two classes of solutions to minimize spacecraft charging:

- To prevent differential charging assure to the extent possible that the entire surface is of uniform conductivity. (Covering solar cells with a transparent but conducting film of Indium oxide is an example).
- Prevent establishment of large potentials by installing plasma generating devices or plasma jets which can actively balance currents to spacecraft surfaces. The devices are often called “plasma contactors” and they effectively “ground” the spacecraft to the surrounding plasma by emitting beams of electrons or ions.

Generally, the above solutions have only been adopted on NASA science mission, operational, DoD, mission and commercial satellites have not generally implemented those solutions. Instead, a set of engineering solutions have been adopted. Most notably, cables should not be run outside the satellite body except where absolutely necessary. Multiple ground loops must be eliminated. Critical command circuitry is designed to not trigger on single pulses.

B Orbital Debris

1 Space Operations and Orbital Debris

We begin our look at this problem with the first chapter from a recently issued report by the Committee on Space Debris of the National Research Council called:

Orbital Debris - A Technical Assessment
National Academy Press, (1995)
2101 Constitution Ave., Washington DC 20418

Space Operations

In the 37 years since the launch of Sputnik 1, space operations have become an integral component of the world's economy, scientific activities, and security systems. Orbital debris is inextricably linked with these operations: debris is created in the course of these operations and is important because it poses a potential threat to future operations. Understanding some of the characteristics of historical, current, and future space operations is thus essential to understanding the overall debris problem.

Today, spacecraft owned by 23 nations and several international organizations (representing more than 100 countries) support a wide variety of important missions, including communications, navigation, meteorology, geodesy and geophysics, remote sensing, search and rescue, materials and life sciences, astrophysics, and national security. A broad spectrum of simple and sophisticated functional spacecraft, with masses ranging from tens of kilograms to tens of metric tons and operational lives ranging from one week to more than ten years are employed to carry out these space activities.

These spacecraft are placed into orbit by a wide variety of launch vehicles. These launch vehicles, which may be either solid or liquid fueled, use multiple stages (some of which may themselves go into orbit) to place spacecraft into their desired orbit. Some spacecraft retain the stage used to perform their orbital insertion maneuver, and most spacecraft have some propulsive capability for attitude control and performing orbital corrections. These spacecraft are placed into orbits from which they can accomplish their particular mission effectively, resulting in a highly non-uniform distribution of spacecraft about the Earth. A few spacecraft each year are launched out of Earth orbit and into interplanetary space; the hazard to future space operations from these probes is utterly negligible.

The distribution of spacecraft around the Earth at the start of 1994 is displayed in Figure 8-5. This distribution is not static; as missions, technologies, and available launch vehicles change, the distribution of functional spacecraft also changes. For example, over the past three decades, the annual percentage of new space missions to orbits above LEO has been increasing; in 1993, High Earth Orbits (HEOs) were the final destination of 42 percent of successful launchings worldwide. Proposed future constellations of communications spacecraft in LEO may reverse this trend.

Figure 8.6 reveals a few characteristics of the current spacecraft population. Most spacecraft reside in LEO, but there are three significant concentrations in higher orbits. There is a concentration of spacecraft (performing Earth observation and communications missions) in

GEO. A second concentration in and near circular semisynchronous orbits is made up of spacecraft from the U.S. Global Positioning System (GPS) and the Russian Global Navigation Satellite System. There is also a significant population of spacecraft in highly elliptical Molniya-type orbits, including Commonwealth of Independent States (CIS) early warning and communications constellations. (In this report, we will refer to pre-1992 space activities of the former USSR as “Soviet” and those of 1992 and later as either “Russian” or of the CIS, as appropriate.) In LEO, notable peaks exist around 1,400 to 1,500 km (due in part to a constellation of Russian communication spacecraft and debris from three breakups of Delta rocket bodies) and 900 to 1,000 km (due in part to Sun-synchronous remote sensing and navigation spacecraft and their associated debris).

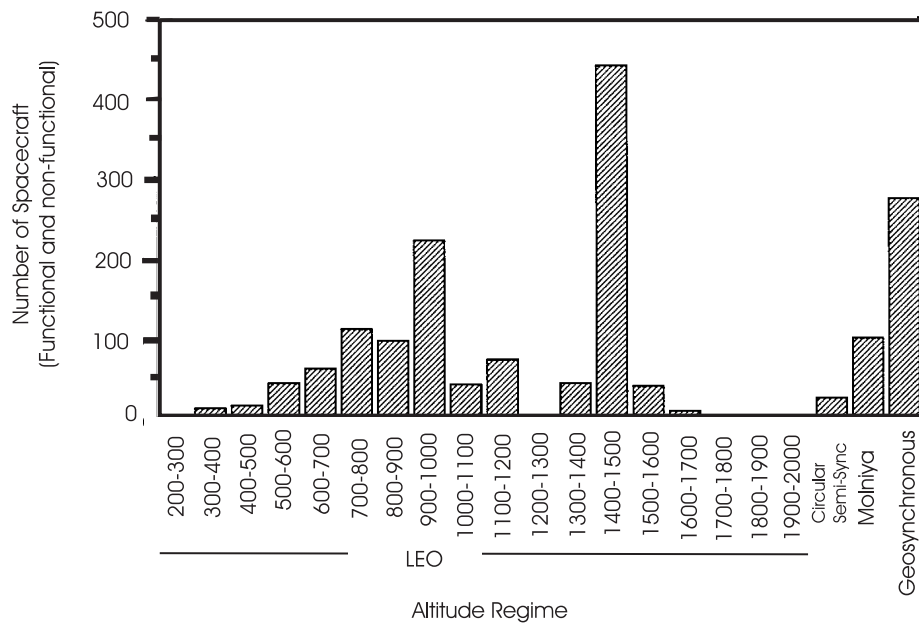


Figure 8.6. Spacecraft population in Earth orbit, 1994.

Most space activities involving humans occur below about 600 km; there are currently few spacecraft in these low orbits because atmospheric drag at these altitudes causes fairly rapid orbital decay.

Table 8.2
Debris Size Conventions - This report uses three general size ranges to categorize debris

Size Category	Approximate Diameter	Approximate Mass	Estimated Number	Detectability	Probable Damage to Spacecraft
Large	>10 cm	>1 kg	8,000	May be catalogable in LEO	Probable loss of spacecraft and possible catastrophic breakup
Medium	1 mm - 10 cm	1 mg - 1 kg	$10^7 - 10^8$	Too small to catalog, too few for most in situ sampling	Ranges from surface degradation through component damage and loss of spacecraft capability
Small	< 1 mm	< 1 mg	10^{12}	Detectable by in situ sampling	Degradation of surface and possible damage to unprotected components

Types of Orbital Debris

The more than 3,600 space missions since 1957 have left a legacy of thousands of large and perhaps tens of millions of medium-sized debris objects in near-Earth space. Unlike meteoroids, which pass through and leave the near-Earth area, artificial space debris orbit the Earth and may remain in orbit for long periods of time. Of the 23,000 space objects officially cataloged by the U.S. Space Surveillance Network (SSN) since the beginning of space age, nearly one-third remain in orbit about the Earth; the majority of these are expected to stay in orbit for tens or hundreds of years. The increasing population of cataloged space objects is represented in Figure 8.7. It is imperative to note that this figure shows only the objects large enough to be repeatedly tracked by ground-based radar. The vast majority of debris is too small to be tracked and is not represented.

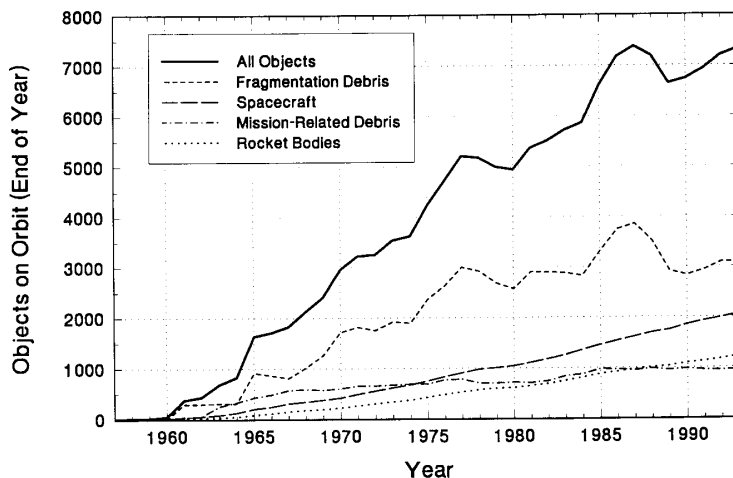


Figure 8.7. On-Orbit cataloged population (corrected for delayed cataloging).

Objects in Earth orbit that are not functional spacecraft are considered debris. Spacecraft that are passive, serving as platforms for laser ranging experiments, atmospheric density measurements, or calibration of space surveillance sensors are considered functional, as are spacecraft that are currently in a reserve or standby status and may be reactivated in the future to continue their mission. Each other type of object in Earth orbit may be classified as belonging to one of four types of debris: nonfunctional spacecraft, rocket bodies, mission-related debris, and fragmentation debris. Figure 8.8 indicates the relative numbers of cataloged functional spacecraft and debris as of March 1994. More than 90% of these cataloged space objects are of U.S. or Soviet/CIS origin, while the remainder belong to nearly 30 other countries or organizations.

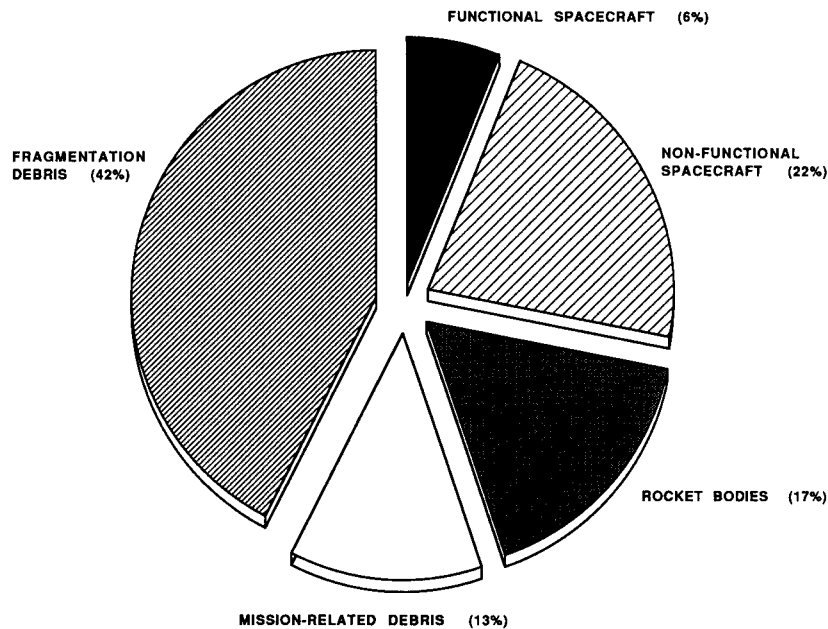


Figure 8.8. Cataloged space objects by category, 1994.

Nonfunctional Spacecraft

Functional spacecraft represent only about one-fifth of the spacecraft population in Earth orbit - the large majority of orbiting spacecraft are nonfunctional. With very few exceptions, functional spacecraft that reach their end of life (EOL), through either termination or malfunction, are left in their former orbit or are transferred to slightly higher or lower altitudes (i.e., are reorbited). Typically, EOL reorbiting maneuvers are performed only by Geosynchronous or semisynchronous spacecraft and by LEO spacecraft carrying nuclear materials. Historically, these EOL maneuvers have almost always resulted in longer orbital lifetimes. Only crewed vehicles and a few other spacecraft (e.g., reconnaissance or space station related) in very low orbits are normally returned to Earth at the conclusion of their missions.

Rocket Bodies

The majority of functional spacecraft are accompanied into Earth orbit by one or more stages (or "rocket bodies") of the vehicles that launched them. Usually only one rocket body is left in orbit for missions to LEO, but the launch vehicle of a high-altitude spacecraft such as GEOS (Geostationary Operational Environmental Satellite) may release up to three separate rocket bodies in different orbits along the way to its final destination. Relatively few spacecraft types (e.g., the U.S. National Oceanic and Atmospheric Administration and Defense Meteorological Satellite Program meteorological spacecraft and Soviet nuclear-powered ocean reconnaissance spacecraft) are designed to retain their orbital insertion stages and leave no independent rocket bodies. Figure 8.9 depicts the rocket bodies and other large debris placed into various orbits in the course of a Proton launch vehicle's delivery of a payload to GEO.

The presence of rocket bodies in orbit is of particular importance to the future evolution of the Earth's debris population due to their characteristically large dimensions and to the potentially explosive residual propellants and other energy sources they may contain. Although the largest stages, which are generally used to deliver spacecraft and any additional stages into LEO, usually reenter the atmosphere rapidly, smaller stages used to transfer spacecraft into higher orbits and insert them into those orbits remain in orbit for long periods of time. Many of these rocket bodies are in orbits that intersect those used by functional spacecraft.

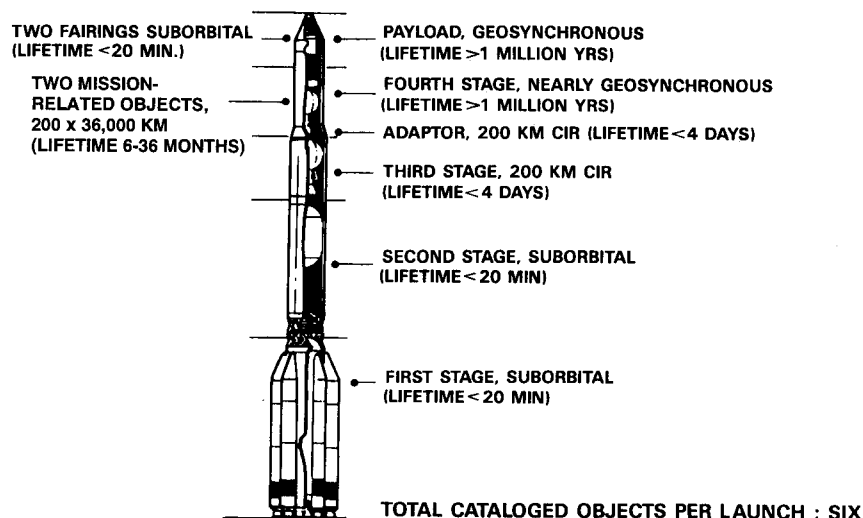


Figure 8.9. Typical debris produced in a Proton launch to GEO.

Mission-Related Debris

Other space objects, referred to as mission-related debris, may be released as a result of a spacecraft's deployment, activation, and operation. Parts of explosive bolts, spring release mechanisms, or spin-up devices may be ejected during the staging and spacecraft separation process. Shortly after entering orbit, the spacecraft may release cords securing solar panels and other appendages or eject protective coverings from payload and attitude control sensors. The amount of debris released can be quite large; a detailed study of the debris released by one Russian launch mission revealed that 76 separate objects were released into orbit from either the launch vehicle or the spacecraft. Numerous debris may also be created during a spacecraft's active life. For example, during the first four years of its operation, more than 200 pieces of mission-related debris linked with the Mir space station were cataloged. Although the occasional item accidentally dropped by a cosmonaut or astronaut may be newsworthy, the majority of this type of debris is intentionally dumped refuse. Since mission-related debris are often relatively small, only the larger items can be detected and cataloged by present-day ground-based surveillance networks.

Another type of mission-related debris comes from the operation of solid rocket motors normally used as final transfer stages, particularly on GEO missions. Current solid rocket fuel usually employs significant quantities of aluminum mixed with the propellant to dampen burn rate instabilities. However, during the burning process, large numbers of aluminum oxide (Al_2O_3) particles are formed and ejected through a wide range of flight path angles at velocities up to 4 km/s. These particles are generally believed to be no larger than 10 microns in diameter, but as many as 10^{20} may be created during the firing of a single solid rocket motor, depending on the distribution of sizes produced. While the orbital lifetimes of individual particles are relatively short, a considerable average population is suggested by examinations of impacts on exposed spacecraft surfaces. More than 25 solid rocket motor firings were conducted in orbit during 1993.

More recently, attention has been drawn to another side effect of solid rocket motors. Ground tests indicate that in addition to the large number of small particles, a smaller number of 1-cm or larger lumps of Al_2O_3 are also ejected during nominal burns. The only indication of the existence of such objects are data from ground tests carried out at Marshall Space Flight Center, Alabama, and the Arnold Engineering and Development Center (Siebold et al., 1993). These medium-sized particles, which have lower characteristic ejection velocities and smaller area-to-mass ratios than the smaller particles, may also be longer-lived than the small particles and could pose a long-term hazard to other Earth-orbiting space objects.

Fragmentation Debris

Fragmentation debris - the single largest element of the cataloged Earth-orbiting space object population - consists of space objects created during breakups and the products of deterioration. Breakups are typically destructive events that generate numerous smaller objects with a wide range of initial velocities. Breakups may be accidental (e.g., due to a propulsion system malfunction) or the result of intentional actions (e.g., space weapons tests). They may be caused by internal explosions or by an unplanned or deliberate collision with another orbiting object.

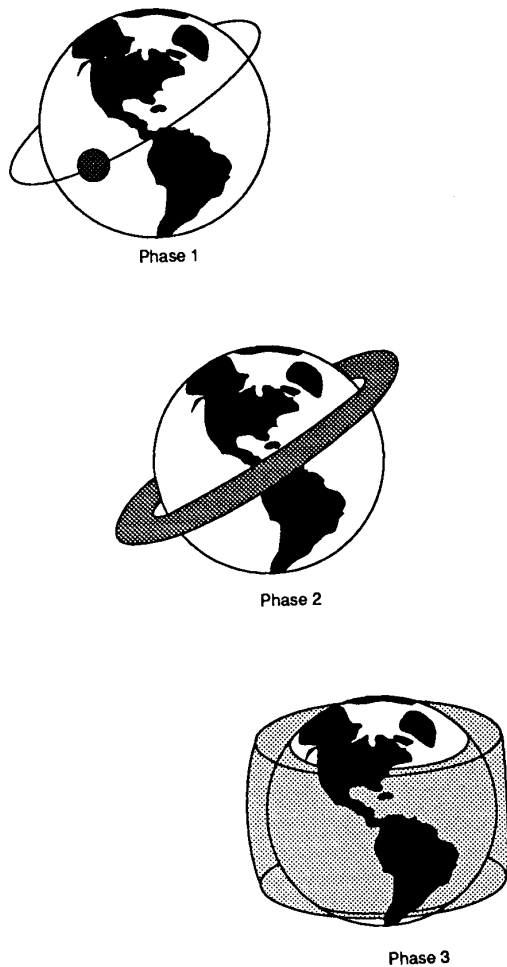


Figure 8.10. Evolution of a debris cloud.

In contrast, debris fragments that are the product of deterioration usually separate at low relative velocity from a spacecraft or rocket body that remains essentially intact. Products of deterioration large enough to be detected from Earth are occasionally seen - probably such items as thermal blankets, protective shields, or solar panels. Most such deterioration is believed to be the result of harsh environment factors, such as atomic oxygen, radiation, and thermal cycling. During 1993 the still-functional COBE (Cosmic Background Explorer) spacecraft released at least 40 objects detectable from Earth - possibly debonded thermal blanket segments - in a nine-month period, perhaps as a result of thermal shock.

Another serious degradation problem involves the flaking of small paint chips as a space object ages under the influence of solar radiation, atomic oxygen, and other forces. Paint, which is used extensively on both spacecraft and rocket bodies for thermal control reasons, can deteriorate severely in space, sometimes in a matter of only a few years. The potential magnitude of the problem was not fully recognized until the 1983 flight of the STS-7 Space Shuttle mission, when an impact crater on an orbiter window was apparently caused by a paint chip smaller than a millimeter in diameter. Subsequent analyses of spacecraft components returned from LEO have

Since 1961, more than 120 known breakups have resulted in approximately 8,100 cataloged items of fragmentation debris, more than 3,100 of which remain in orbit. Fragmentation debris thus currently makes up more than 40 percent of the U.S. space object catalog (and undoubtedly represents an even larger fraction of non-cataloged objects). The most intensive breakup on record was the 1987 breakup of the Soviet Kosmos 1813, which generated approximately 850 fragments detectable from the Earth. The fragmentation debris released from a breakup will be ejected at a variety of initial velocities. As a result of their varying velocities, the fragments will spread out into a toroidal cloud that will eventually expand until it is bounded only by the limits of the maximum inclinations and altitudes of the debris. This process is illustrated in Figure 8-10. The rate at which the toroidal cloud evolves depends on both the original spacecraft's orbital characteristics and the velocity imparted to the fragments; in general, the greater the spread of the initial velocity of the fragments, the faster will the evolution occur.

confirmed the presence of a large population of paint particles, even though the orbits of individual particles decay quite rapidly.

Perturbation Forces affecting Space Objects

Once in orbit, debris is affected by perturbing forces that can alter its trajectory and even remove it completely from orbit. Other than the gravitational attraction of the Earth, the primary forces acting on a space object in lower orbits (below about 800 km) are atmospheric drag and gravitational perturbations from the Earth. These gravitational perturbations, however, although affecting some orbital parameters, do not generally strongly affect orbital lifetime. For space objects in higher orbits, solar and lunar gravitational influences become more important factors. Small debris can also be affected by solar radiation pressure, plasma drag, and electrodynamics forces, although the effects of plasma drag and electrodynamics forces are typically dwarfed by the effects of solar radiation pressure.

The rate at which a space object loses altitude is a function of its mass, its average cross-sectional area impinging on the atmosphere, and the atmospheric density. Although the Earth's atmosphere technically extends to great heights, its retarding effect on space objects falls off rapidly with increasing altitude. Atmospheric density at a given altitude however, is not constant and can vary significantly (particularly at less than 1,000 km) due to atmospheric heating associated with the 11-year solar cycle. This natural phenomenon has the effect of accelerating the orbital decay of debris during periods of solar maximum (increased sun-spot activity and energy emissions). During the last two peaks in the solar cycle, the total cataloged space object population actually declined, because the rate of orbital decay exceeded the rate of space object generation via new launches and fragmentations.

Figure 8.11, which displays the predicted orbital lifetimes for a number of different objects in circular LEOs at different periods in the solar cycle, illustrates the importance of cross-sectional-area-to-mass ratio, altitude, and solar activity in determining orbital lifetimes in LEO. First, objects with low ratios of cross-sectional area to mass decay much more slowly than objects with high area- to-mass ratios. Second, objects at low altitude experience more rapid orbital decay than objects at high altitude. Finally, objects decay much more rapidly during periods of solar maximum than during the solar minimum.

The combination of all of these forces has caused approximately 16,000 cataloged objects to reenter the atmosphere since the beginning of the space era. In recent years, an average of two to three space objects large enough to be cataloged (as well as numerous smaller debris particles) reenter the Earth's atmosphere each day. Over the course of a year, this amounts to hundreds of metric tons of material. This material is composed primarily of large objects that were launched into low orbits (most of the mass is in the form of large multiton rocket bodies) and small objects with high cross-sectional-area-to-mass ratios. Seldom do any larger objects initially placed into orbits higher than 600 km reenter the atmosphere.

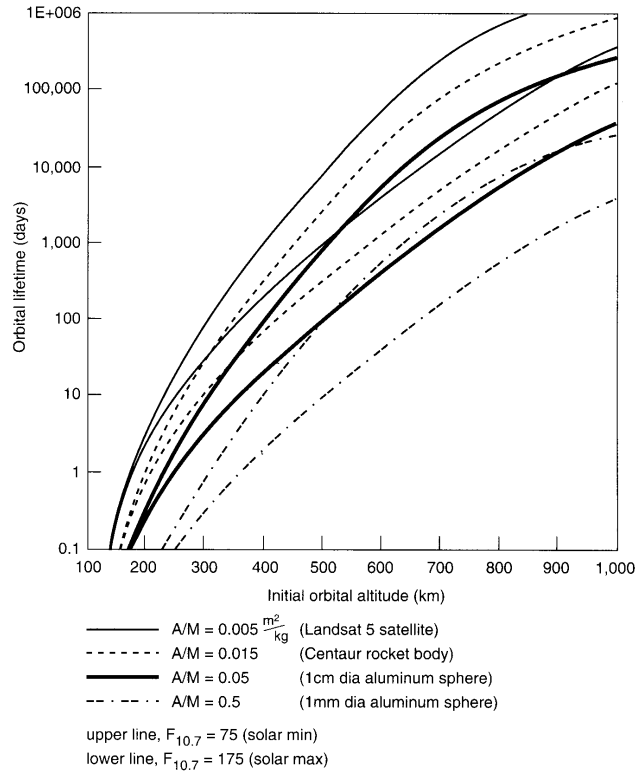


Figure 8.11. Orbital decay time versus altitude

2 Hypervelocity Impacts

The kinetic energy of particles moving at speeds of 10-15 km/sec is roughly equivalent to the energy released by the explosion of 40 times its mass of TNT. For example, a 1 cm diameter aluminum sphere (whose mass is about 1.4 gr.) moving at 13 km/sec has a kinetic energy equivalent to the energy released by the explosion of 56gr. of TNT (about 0.24 MJ). The characteristic of a hypervelocity collision depends on the initial speed as shown.

speed v (km/sec)	
$v < 2$	Projectile remains intact
$2 < v < 7$	Projectile shatters into fragments
$7 < v < 11$	Projectile melts upon impact
$v > 11$	Projectile may vaporize

In addition some fraction of the available energy will be dissipated in generating a crater in the target material. If the crater is sufficiently deep the target material may be penetrated as indicated by the following empirical relationship for the penetration depth t (in cm).

$$t(\text{cm}) = k M_p^{\alpha/3} \rho_t^{\beta/3} v_{\perp}^{\gamma/3} \quad (\text{Eqn. 8.10})$$

M_p = mass of projectile (in gr.)

ρ_t = density of target material (in gr./cm³)

v_{\perp} = Projectile velocity component perpendicular to the surface (in km/sec)

k = material specific constant

The constants α , β and γ are empirically determined and on the order of $\alpha \sim 1$, $\beta \sim \frac{1}{2}$ and $\gamma \sim 2$.

Results from the Long Duration Exposure Facility (LDEF) have yielded the following expressions for the penetration depth in aluminum.

$$t(\text{cm}) = 0.72 M_p^{0.352} \rho_t^{1/6} v_{\perp}^{0.875} \quad (\text{Eqn. 8.11})$$

There also exists an approximate relation for crater depth when the target material is not penetrated. In that case the crater depth P (cm) is

$$t(\text{cm}) = 0.42 M_p^{0.352} \rho_t^{1/6} v_{\perp}^{2/3} \quad (\text{Eqn. 8.12})$$

Both of the above relations are illustrated in Fig 8-12 below for an aluminum target surface and a 10 km/sec incident projectile.

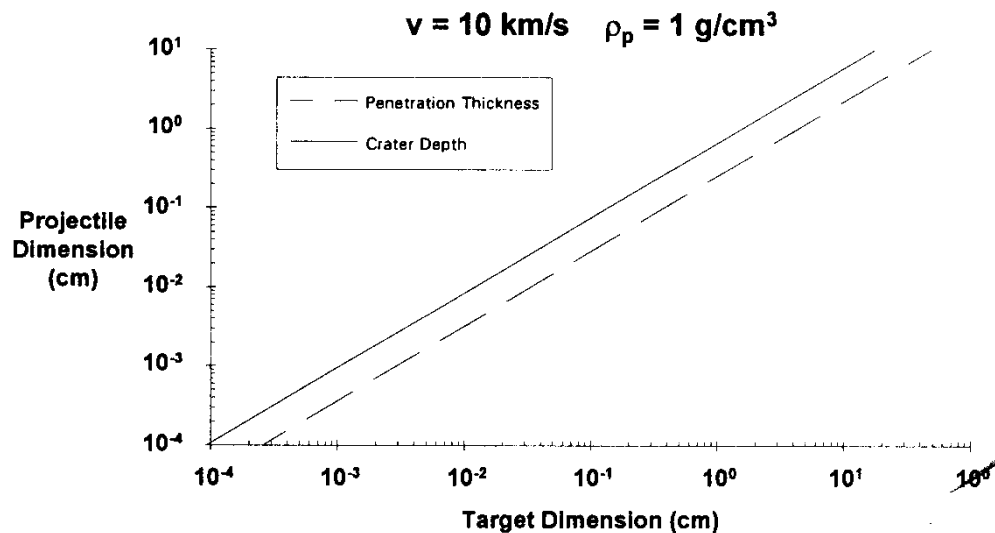


Figure 8.12. Crater diameter and penetration thickness in aluminum

As was shown in the previous section the effects of orbital debris collision range from surface degradation all the way to catastrophic destruction of the spacecraft. For a given projectile dimension (and speed) the crater formed in a block of material is less deep than the thickness of a flat plate penetrated. This effect is due to spallation and shockwave reflection from the back surface of the plate.

3 Impact Probabilities.

A number of complex models have been constructed to predict the flux of orbital debris objects as a function of debris mass (or size) and altitude. The results of one such projection is shown in Figure 8-13.

To calculate the impact probability we must first find N the number of impacts in time T which is given by

$$N = \int_t^{t+T} FA \, dt \quad (\text{Eqn. 8.13})$$

where F = Flux (Particles/m²/year are the usual units); A = Surface area (m²)

Once we have N we can obtain the probability of n impacts from the well-known Poisson distribution as:

$$P_n = \frac{N^n}{n!} e^{-N} \quad (\text{Eqn. 8.14})$$

To estimate the number of impacts due to objects greater than a specific mass we exclude the lower mass particles from the flux in calculating N .

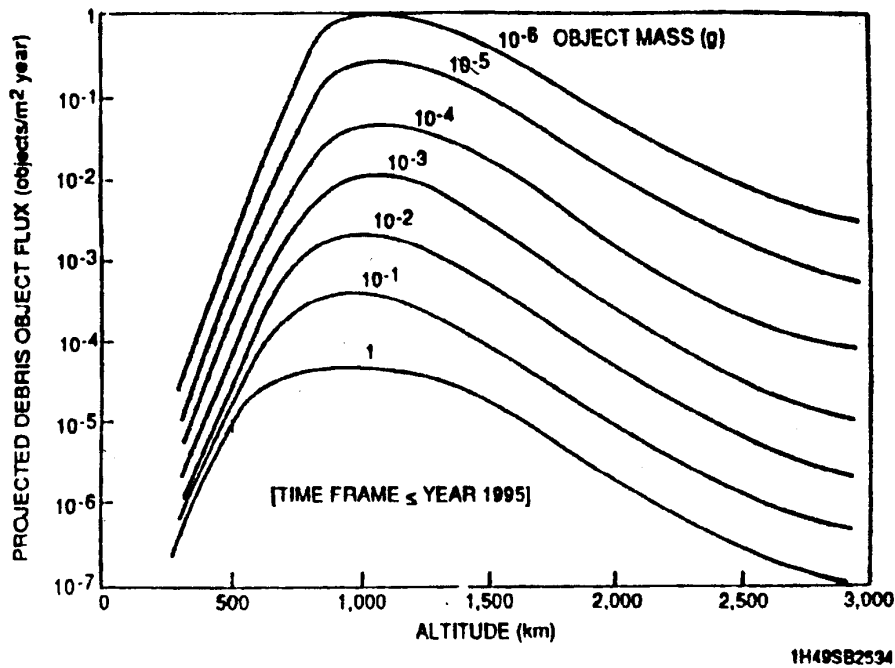


Figure 8.13 1995 Space Debris Projection

Although the Orbital Debris environment is not well defined or understood a number of models have been constructed which approximate the flux F for specified particle diameter, orbital altitude, solar activity index, orbital inclination, year of observation and assumed debris growth and decay rates.

Figure 8.14 shows the results for a typical LEO orbit and specified conditions

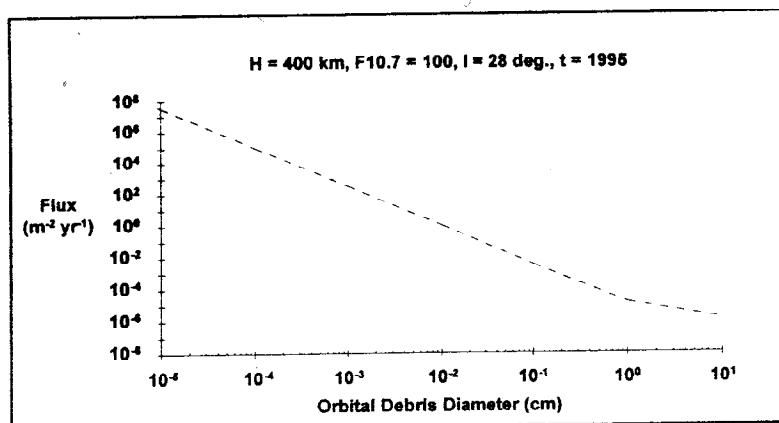


Figure 8.14

Suppose we wish to calculate the probability for 0,1 and 2 impacts of a 1mm projectile on a spacecraft of 5m² surface area in one year

$$N = \int_0^1 FAdt = 5 \times 10^{-2} \quad (\text{Eqn. 8.15})$$

using the Poisson distribution we obtain

$$\begin{aligned} P_0 &= \frac{N^0}{0!} e^{-N} = e^{-N} = e^{-0.05} = 0.9512 \\ P_1 &= \frac{N^1}{1!} e^{-N} = 0.05 e^{-0.05} = 0.0475 \\ P_2 &= \frac{N^2}{2!} e^{-N} = \frac{(0.05)^2}{2} e^{-0.05} = 0.0012 \end{aligned} \quad (\text{Eqn. 8.16})$$

as the probabilities for 0, 1 or 2 impacts/year

4 Design Considerations

Spacecraft thermal blankets or structural panels can provide adequate shielding for small (diameter < .1mm) debris. For larger debris orienting sensitive surfaces away from ram direction or flying at altitudes or orbital inclinations that minimize debris flux can be helpful.

Considerable work on debris “bumpers” is currently underway. These bumpers consist of multiple sheets of material separated by few centimeters. The space between them may be empty or filled with materials such as Kevlar or Nextel. The outer shield is expected to be penetrated by the incident particle, but the impact would fragment the projectile into many smaller pieces which would then be stopped by the back layer of the bumper.

C Surface Effects

1 Atomic Oxygen Effects

At 300 km altitude ambient atmospheric density is about ten orders of magnitude below that encountered at sea level. However there are still $\sim 10^{15}$ oxygen atoms/m³ and 1m² spacecraft orbiting at 8 km/sec will undergo about 10^{19} collisions with ambient atoms/sec with an effective collision energy of about 5eV. Since atomic oxygen is highly reactive these collisions result in oxidation and erosion of surface materials.

Consider the erosion of material from a surface due to a flux of oxygen atoms. The mass loss in time dt is given by

$$dm = \rho(RE) \phi A dt \quad (\text{Eqn. 8.17})$$

where: ρ (gr./cm³) is the material density

ϕ (OA atoms/cm²-sec) is the atomic oxygen flux

(RE) is the reaction efficiency in units (cm³/OA atom).

The reaction efficiency is an experimentally determined quantity, which may be a function of various parameters such as surface temperature, uv flux, atomic oxygen flux, etc.

The above equation can also be rewritten to show the rate of change of thickness of the target.

$$\frac{dx}{dt} = (RE) \phi \left(\frac{\text{cm}}{\text{sec}} \right) \quad (\text{Eqn. 8.18})$$

This rate of change can be as high as a few tenths of a millimeter per year for certain thermal control materials such as kapton.

Table 8.3 Atomic Oxygen Reaction Efficiency - Reaction Efficiency $\times 10^{-24}$ cm³/atom

Material	Range	Best Value
Aluminum	-	0.00
Carbon	0.9-1.7	-
Epoxy	1.7-2.5	-
Fluoropolymers		
-FEP Kapton	-	0.03
-Kapton F	-	<0.05
-Teflon, FEP	-	<0.05
-Teflon	0.03-0.50	-
Gold	-	0.0
Indium Tin Oxide	-	0.002
Mylar	1.5-3.9	-
Paint		
-S13GLO	-	0.0
-YB71	-	0.0
-Z276	-	0.85
-Z302	-	4.50
-Z306	-	0.85
-Z853	-	0.75
Polyimide		
-Kapton	1.4-2.5	-
-Kapton H	-	3.04
Silicones		
-RTV560	-	0.443
-RTV670	-	0.0
Silver	-	10.5
Tedlar		
-Clear	1.3-3.2	-
-White	0.05-0.6	-

It should also be kept in mind that significant changes in the thermal and optical properties of surfaces can occur without appreciable mass loss.

Another very important effect is the deterioration of interconnects between solar cells which are often made of silver. If unprotected these connecting wires may literally be worn away and the solar array may fail.

2 UV Degradation

Essentially all solar radiation below 0.3mm in wavelength is absorbed by the atmosphere before it reaches the surface of the earth. A satellite on the other hand is exposed to the full spectrum of solar radiations including UV and X Rays. The photon energy corresponding to $\lambda = 0.3 \mu\text{m}$ is about 4 eV. As the table shows typical UV photon energies are sufficient to break many chemical bonds and thereby alter the physical properties of surfaces.

Table 8.4 - Chemical Bond Energy

Chemical	Bond	Bond Energy at 25 degrees C (kcal/mole)	Bond Energy at 25 degrees C (eV)	Wavelength (microns)
C-C	Single	80	3.47	0.36
C-N	Single	73	3.17	0.39
C-O	Single	86	3.73	0.33
C-S	Single	65	2.82	0.44
N-N	Single	39	1.69	0.73
O-O	Single	35	1.52	0.82
Si-Si	Single	53	2.30	0.54
S-S	Single	58	2.52	0.49
C-C	Double	145	6.29	0.20
C-N	Double	147	6.38	0.19
C-O	Double	176	7.64	0.16
C-C	Triple	198	8.59	0.14
C-N	Triple	213	9.24	0.13
C-O	Triple	179	7.77	0.16

One of the important quantities for thermal control of spacecraft is the solar absorptance α_s which varies from 0 for perfect reflector to 1 for a perfect absorber. Large changes in the absorptance (on the order of 50%) have been observed due to UV irradiation for certain materials and exposure times on the order of 1000 hrs.

3 Sputtering

When atoms or ions of sufficient energy impact a solid surface they can eject atoms in a process called sputtering. In Table 8.5 are listed the minimum energies which an incident particle must have to cause sputtering on various target materials.

Table 8.5 - Sputtering Thresholds - Bombarding Gas Threshold (eV)

Target Material	O	O ₂	N ₂	Ar	He	H
Ag	12	14	13	17	25	83
Al	23	29	27	31	14	28
Au	19	15	15	15	53	192
C	65	82	79	88	40	36

Cu	15	22	21	24	20	60
Fe	20	28	27	31	23	66
Ni	20	29	27	31	24	72
Si	31	39	37	42	18	40

The energies which are available in collisions between atmospheric atoms and moving satellite surface are generally lower than the sputtering thresholds. Thus collisions due to the orbital motion of the satellite do not cause a significant amount of sputtering.

However we did see earlier that satellite surfaces can become negatively charged up to several kilovolts and these negatively charged surfaces will accelerate any positive ions in the atmosphere toward the surface yielding impact energies of hundreds or even thousands of eV more than enough to cause sputtering as shown in Table 8.6 below for the case of 100eV incident particles on various target surfaces.

Table 8.6 - Sputtering Yields (atoms/particle) at 100eV Impact Energy

Target Material	O	O ₂	N ₂	Ar	He	H
Ag	0.265	0.498	0.438	0.610	0.030	-
Al	0.026	0.076	0.060	0.110	0.020	0.010
Au	0.154	0.266	0.244	0.310	-	-
C	-	-	-	-	0.008	0.008
Cu	0.385	0.530	0.499	0.600	0.053	-
Fe	0.069	0.153	0.129	0.200	0.028	-
Ni	0.120	0.247	0.239	0.270	0.029	-
S	0.029	0.054	0.046	0.070	0.023	0.002

On extended missions the sputtering of materials, in particular metals can cause erosion as well as changes in surface properties.

4 Molecular Contamination

Many materials when placed in a vacuum will “outgas” meaning that atoms or molecules will spontaneously leave the surface. In space outgassing materials may deposit contaminants onto sensitive surfaces such as thermal control panels, solar arrays or optical surfaces thereby altering their thermal or optical properties. To make matters worse we sometimes observe synergistic effects which result in total degradation which is greater than the sum of its parts. One such example is the interaction of solar UV with contamination films deposited by outgassing. By polymerizing the contaminant molecules films can be made to adhere to warm surfaces such as solar panels. An example of this effect is shown in the graph below.

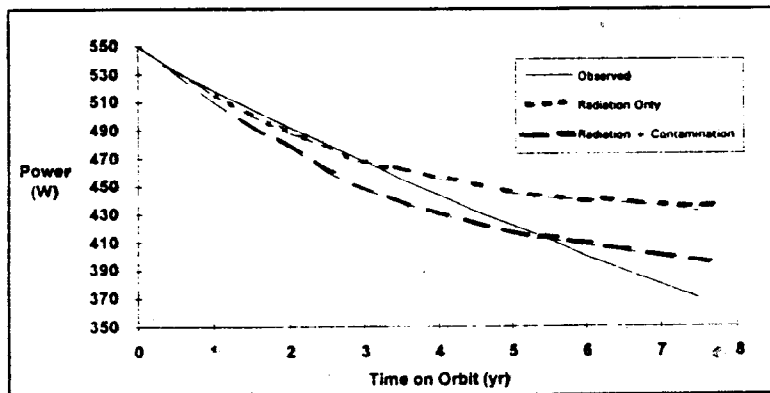


Fig. 2.12 GPS Block I solar power degradation.
(©1991 AIAA—reprinted with permission)

Figure 8.15

D Homework

1. Using the formula for attracted and repelled species, plot the magnitudes (absolute values) of the ion and electron currents to a satellite for the following parameters:

proton plasma: ion mass = 1.67×10^{-27} kg

ion temperature: ~~20 keV~~

electron temperature: ~~20 keV~~ use $T = 5$ KeV for ions and electrons

electron density = 1×10^6 m³

$\Phi = 0$ to -20 kV

note the formula to be using are: $I = I_0 e^{-q\Phi/kT}$ (Eqn. 8.7a) and $I = I_0 \left(1 - \frac{q\Phi}{kT}\right)$ (Eqn. 8.7b).

$$I_0 = q n v; \quad v = \sqrt{\frac{kT}{m}};$$

The charge q will be positive for ions, and negative for electrons

($e = 1.6 \times 10^{-19}$, $m_e = 9.1 \times 10^{-31}$)

Assume the satellite has a surface area of 1 m^2 .

The two curves will intersect. The intersection point is where the net current is zero, and hence the equilibrium point.

Chapter 9 Environmental Effects on Space Systems - Part 2 Radiation Effects

A Passage of ionization through matter.

Ionizing radiation means that enough energy can be transferred to an atomic electron in a single interaction to detach the electron from the atom. The energies required to do this depend on the target atom and range from 5 to 25 eV, usually called the ionization energy. This is for the most loosely bound electron. The inner electrons may require high energies up to several thousand electron volts.

Ordinary matter is very open when viewed on an atomic scale. If we represent a nucleus by a dot 0.1 mm in diameter (just visible) then the next nucleus in a typical solid will be about 10 meters or 30 ft away! So there is plenty of room for ions, electrons or photons to penetrate into the solid. The space between the nuclei is filled with an electron cloud which has little mass but which generates strong electromagnetic fields which really hold the solid together and make it “solid”.

There are 4 classes of particles or “radiations” of interest to us:

- (1) Heavy, charged particles such as protons, α particles or heavier ions
- (2) High energy photons - UV, X Rays, and gamma rays
- (3) Electrons.
- (4) Neutral particles - mostly neutrons.

All of these can cause ionization if they have enough (kinetic) energy.

1 Penetration of heavy charged particles

When a heavy charged particle such as a proton or an α particle (He^{++}) enters a solid it interacts almost entirely with the distributed electrons each of which has a much smaller mass than the projectile:

$$\frac{\text{Proton mass}}{\text{Electron mass}} \cong 1835 \qquad \frac{\alpha \text{ particle mass}}{\text{Electron mass}} \cong 7340$$

We know from mechanics that in collisions of heavy projectiles with a light target only a small fraction of the projectile energy can be transferred to the target:

$$\frac{\Delta E}{E} = \frac{(4)(\text{mass of target})}{\text{Mass of projectile}}$$
 is maximum fractional energy transfer.

Thus as the proton travels through the electron cloud it makes many collisions losing a small fraction of its kinetic energy in each collision until it finally stops. Thus a heavy charged particle has well defined range which represents the distance beyond which no particle in an incident mono-energetic beam will penetrate as shown in Figure 9.1

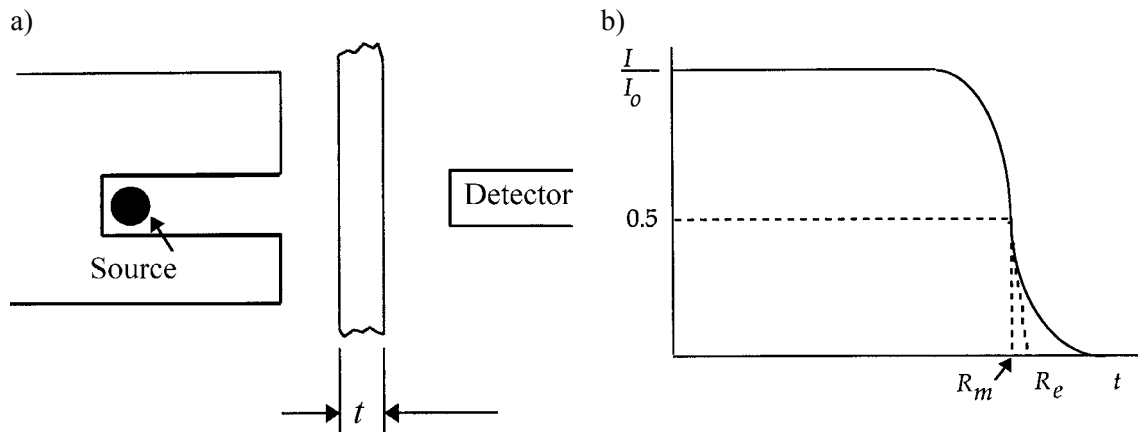


Figure 9.1. An alpha particle transmission experiment. I is the detected number of alpha particle through an absorber thickness, t , whereas I_0 is the number detected without the absorber. The mean range R_m and extrapolated range R_e are indicated.

Another important concept is the stopping power usually written as $-\frac{dE}{dx}$ and shown schematically in Figure 9.2. It is also called the Specific Energy loss.

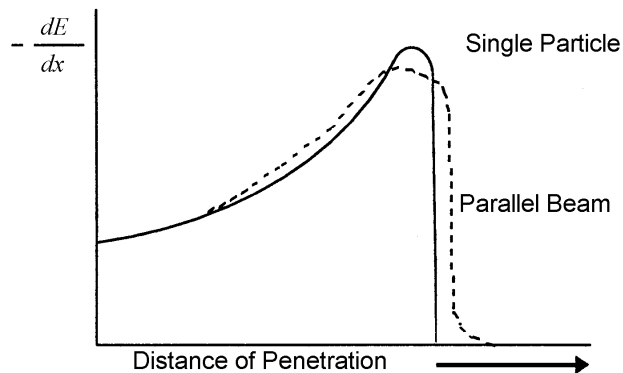


Figure 9.2. The stopping power along an alpha particle track.

As the notation implies $-\frac{dE}{dx}$ represents the rate of energy loss with respect to distance traveled along the trajectory. $-\frac{dE}{dx}$ is small at first when the incident particle velocity is high and then rises to a peak near the end of the range and drops abruptly to 0.

The relation between range R and $-\frac{dE}{dx}$ is fairly simple.

$$R = \int_0^{E_0} \frac{dE}{\left(\frac{dE}{dx}\right)} \text{ since } \frac{1}{\left(\frac{dE}{dx}\right)} = \text{the distance traveled for a loss of unit energy.}$$
 The units are almost as peculiar as the dimensions, again due to the nature of the ultimate applications:

It would be most logical to specify range in centimeters and $\frac{dE}{dx}$ in energy/distance such as MeV/cm. However, it often turns out to be more useful to specify range in terms of the amount of mass of target material penetrated. Hence:

$$R\left(\frac{\text{gr}}{\text{cm}^3}\right) = \rho R(\text{cm})$$

where ρ = density of the target material in $\frac{\text{gr}}{\text{cm}^3}$. Similarly,

$$\frac{dE}{dx}\left(\frac{\text{MeV}}{\frac{\text{gr}}{\text{cm}^3}}\right) = \frac{1}{\rho} \frac{dE}{dx}\left(\frac{\text{MeV}}{\text{cm}}\right)$$

The reason for these units is the fact that R and $\frac{dE}{dx}$ expressed in $\frac{\text{gr}}{\text{cm}^2}$ and $\frac{\text{MeV cm}^2}{\text{gr}}$ respectively are approximately the same for elements close to each other in the periodic table, such as Si and Al. Figure 9.3 shows both the range and the stopping power $-\frac{dE}{dx}$ for protons in silicon, a material of great interest in space applications.

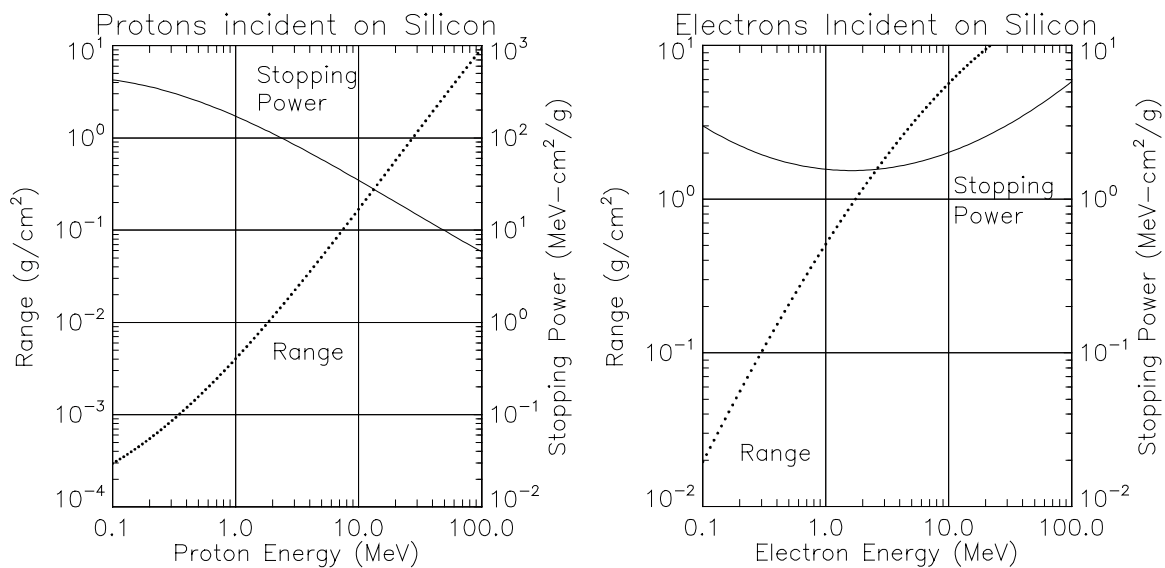


Figure 9.3. Stopping Power and Range Curves for Protons and electrons in Silicon

Another material of considerable importance in spacecraft construction is aluminum.

Figure 9.4a gives the range-energy curve for protons in aluminum, water and lead for protons; Figure 9.4 b shows ranges for electrons. Note that the shapes of the curves are similar - the scales have simply shifted. Typical spacecraft skins are about 1/8 inch (0.3 cm) or 0.86 gm/cm². From the curve in Figure 9.4, this would correspond to the range of approx. a 30 MeV proton.

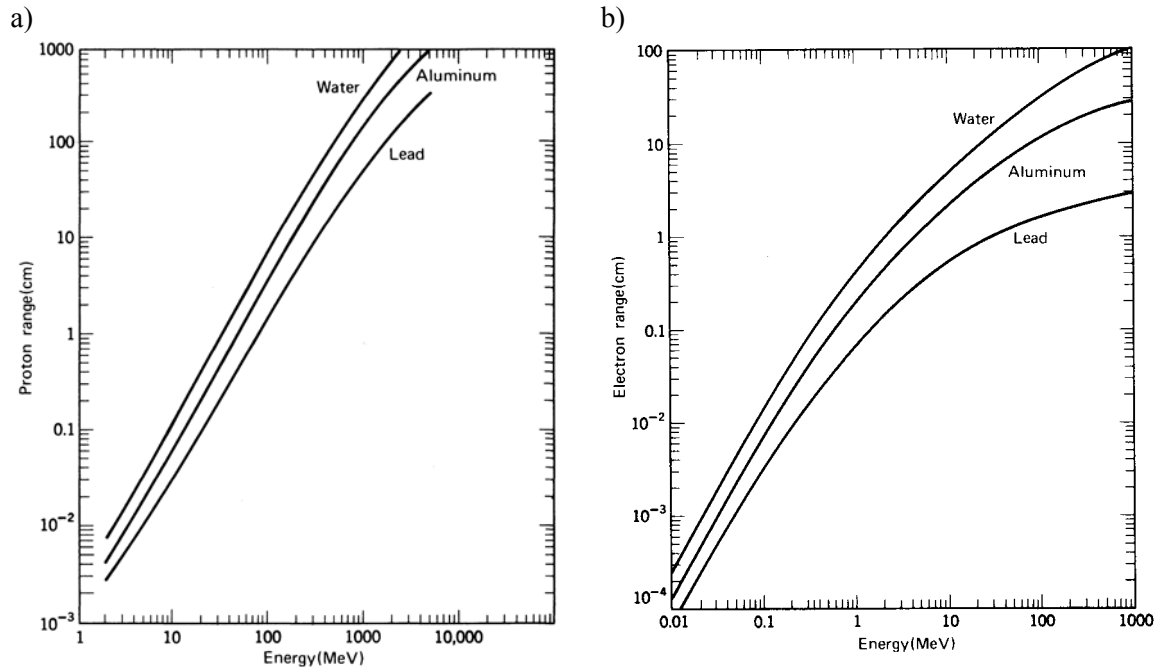


Figure 9.4 Range-Energy Relation for protons in Al for proton energies 0-10⁴ MeV

2 Interaction of Fast Electrons with Matter

Compared to heavy charged particles electrons lose energy at a much lower rate and the path of the fast electron is usually convoluted rather than an essentially straight line path. Large deflection in the incident electron's path are possible because the masses of the projectile and target are equal and therefore a much larger fraction of the incident energy can be lost in a single encounter.

For electrons there also exists a second mechanism by which energy is lost, namely radiation of electromagnetic energy when the incident electron undergoes a sudden acceleration e.g. a collision. This process is usually called Bremsstrahlung (Braking Radiation).

Thus, the total $\frac{dE}{dx}$ for the incident electron has two parts:

$$\frac{dE}{dx} = \left(\frac{dE}{dx} \right)_{\text{coll}} + \left(\frac{dE}{dx} \right)_{\text{rad}}$$

There exists a rule of thumb to estimate the relative magnitude of these two contributions

$$\frac{\left(\frac{dE}{dx} \right)_{\text{rad}}}{\left(\frac{dE}{dx} \right)_{\text{coll}}} = \frac{EZ}{700} \quad \text{E = the incident energy in MeV and Z is the atomic number of the target}$$

For electrons of a few MeV and targets such as AL(13) or Si(14) the above ratio is small showing that collisional energy loss dominates.

Electron Range and Transmission Curves

If we repeat the absorber experiment described above in Figure 9.1 using a source of mono-energetic fast electrons we get a curve like the one shown in Figure 9.5 below.

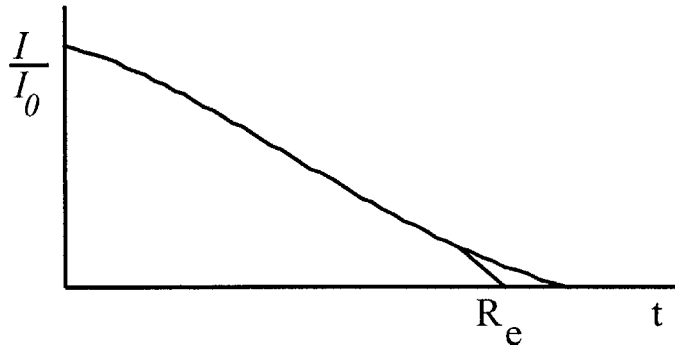


Figure 9.5. Transmission curve for mono-energetic electrons R_e is the extrapolated range.

As can be seen from the curve even for a very thin absorber some electrons are lost because they are scattered out of the beam and do not reach the detector. Thus the plot of the detected number of electrons vs absorber thickness begins to drop immediately and gradually approaches zero for large absorber thickness.

The concept of range is less definite for fast electrons than heavy charged particles. Normally the electron range is taken from a plot such as Figure 9.5 by extrapolation of the linear portion of the curve to zero and represents the absorber thickness required to assure that almost no electrons can penetrate the entire thickness

For equivalent energy the specific energy loss $\frac{dE}{dx}$ is much lower for electrons than that of heavy charged particles so their actual path length in typical absorbers is hundreds of times greater. As a rough rule of thumb electron ranges are about 2mm/MeV for the materials of interest to us (Si and Al) See Figure 9.3 above and 9.6 below

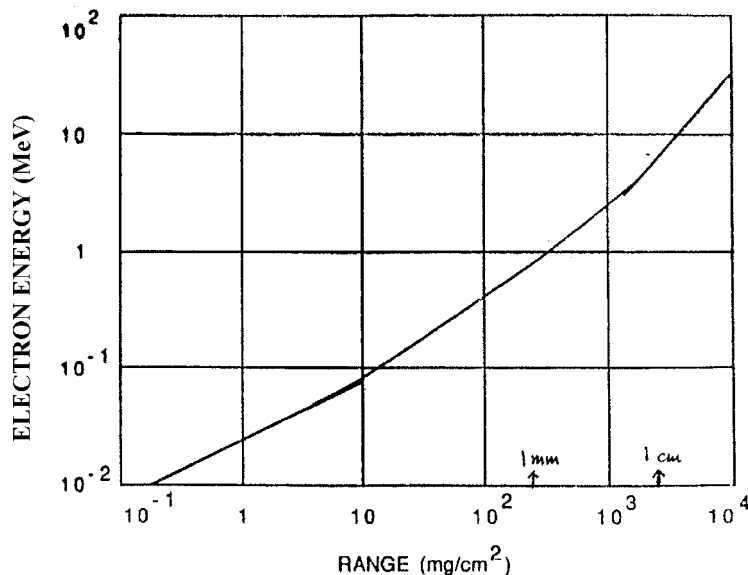


Figure 9.6. Electron Energy Versus Range in Aluminum

3 Interaction of energetic photons with matter

The process by which photons interact with matter is fundamentally different from that of charged particles. The photon will transfer all of its energy in a single interaction and vanish in the process. This leads to an exponential decrease of intensity I (photons/area/sec.) with penetration depth into the absorber. Thus intensity at depth t is $I(t) = I_0 e^{-\mu t}$ where μ is called the linear attenuation coefficient and has units of cm^{-1} . This type of attenuation illustrated in Figure 9.7.

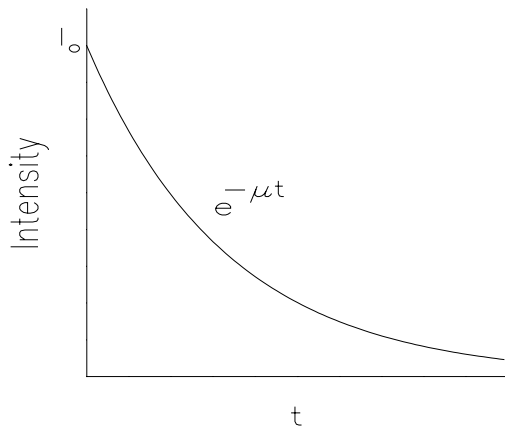


Figure 9.7. The exponential transmission curve for gamma rays measured under “good geometry” conditions.

It again turns out to be convenient to define a mass attenuation coefficient

$$\mu_m = \frac{\mu}{\rho} \left(\frac{\text{cm}^2}{\text{gr}} \right) \text{ and write the depth in the units of } t_m = \rho t \left(\frac{\text{gr}}{\text{cm}^2} \right) \text{ which yields}$$

$$I(t) = I_0 e^{-(\mu/\rho)(\rho t)} = I_0 e^{-\mu_m t_m}$$

Where μ (and μ_m) are functions of the target material and the energy of the incident photons. Figure 9.8 shows μ_m for various materials in the keV and MeV energy range.

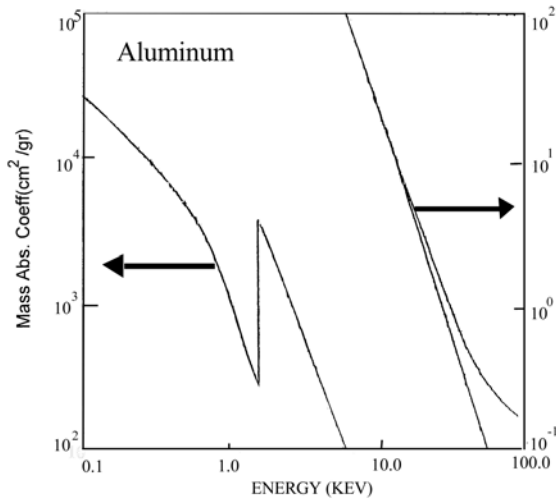


Figure 9.8 a Mass absorption coefficient for X-Rays in aluminum

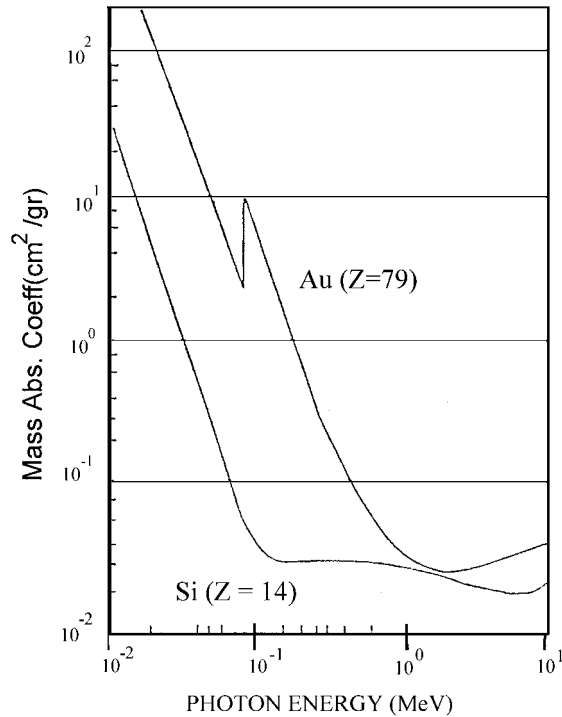


Figure 9.8 b Figure 9.6: Mass Absorption Coefficients for Si, Au for X-Rays of various energies

4 Neutrons

Finally a word about neutral particles, particularly neutrons. Since they carry no charge their primary interaction is with the small but massive nuclei. When such a collision occurs the neutron may be swallowed up by the target nucleus or at the very least suffer a big change in energy. Thus the intensity of a collimated neutron beam decays exponentially as

$$I(t) = I(0)e^{-(\Sigma_{\text{total}})t},$$

where $\Sigma_{\text{total}} = \Sigma_{\text{scatter}} + \Sigma_{\text{absorb}} + \dots$

Neutrons are rare in the natural space environment, but are important in the vicinity of nuclear reactors or nuclear detonations.

B Radiation Units

We are interested in the effects of radiation in matter including biological tissue. We shall here only describe the four basic quantities which are used to do this (They are often loosely and incorrectly applied).

(a) The curie (ci) is a measure of the activity of a radioactive source and is defined as $1 \text{ ci} = 1 \text{ curie} = 3.70 \times 10^{10} \text{ disintegrations/sec}$. This says nothing about the nature of decays. This unit is generally used to describe radioactive sources or materials.

(b) The Roentgen [R] is a measure of exposure. Specifically one Roentgen is defined as the exposure which would deliver 8.78 mJ to 1 kg of dry air. A proper use would be to say: This dental X-Ray beam provided an exposure of 300 mR/sec. Although we use dry air as a measurement substance the exposure itself is independent of materials.

(c) The RAD (from Radiation Absorbed Dose) is a measure of the dose actually absorbed by a specific object in terms of the energy transferred to it. An object is said to have received and absorbed dose of 1 RAD when 10mJ/kg have been deposited in it by ionizing radiation. A proper statement using the concept of absorbed dose would be:

“A whole body short term gamma ray dose of 500 RAD will cause death in 50% of the population exposed to it.”

The SI unit for absorbed dose is the Gray defined by
 $1 \text{ Gray (Gy)} = 100 \text{ RAD} = 1(\text{J/kg})$

(d) The REM stands for (Roentgen Equivalent in Man) and takes account of the fact although different radiations may deliver the same energy/pr unit mass to a body the biological effects can be very different because of the different dE/dx for different types of radiation.

To account for this we multiply the absorbed dose in RAD's by a factor called the Relative Biological Effectiveness or (RBE) factor. For X-Rays and electrons RBE~1, for slow neutrons ~5 and for α particles 10-20, etc.

Thus $(\text{REM}) = (\text{RAD})(\text{RBE})$ converts RADs into REMs.

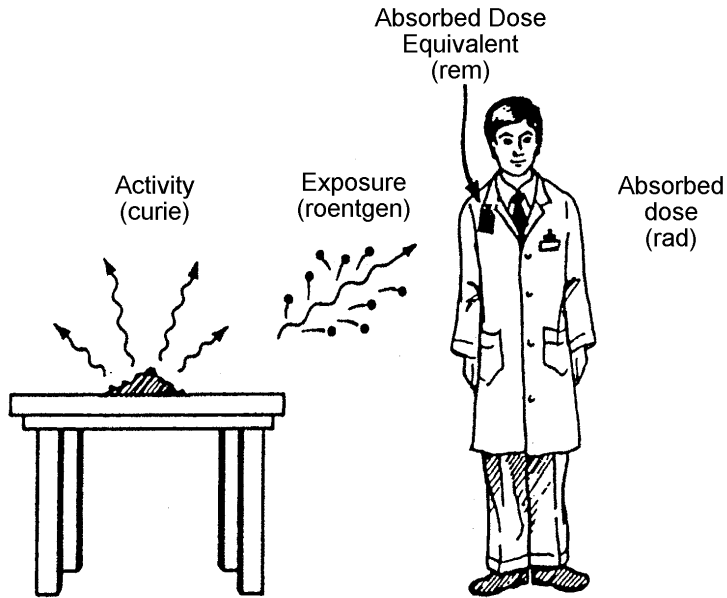


Figure 9.9. (from: Fundamentals of Physics by Halliday & Resnick: Wiley & Sons, 1988)
This sketch should help to clarify the distinction between the curie, the roentgen, the RAD, and the REM.

Table 1 - Acute Dose Effects on Humans (Whole Body Exposure)

Dose (RADs)	Probable Effect
0-50	No obvious effects, blood changes
80-120	10% chance of vomiting/nausea for 1 day
130-170	25% chance of nausea, other symptoms
180-220	50% chance of nausea, other symptoms
270-330	20% deaths in 2-6 weeks, or 3 month recovery
400-500	50% deaths in 1 month, or 6 month recovery
550-750	Nausea within 4 hours, few survivors
1000	Nausea in 1-2 hours, no survivors
5000	Immediate incapacitation, death within 1 week

C Effects of Space Radiation on Systems

Space contains large numbers of high speed atomic particles and energetic photons. Any object placed in space will be impacted by these particles and photons and depending on circumstances varying degrees of radiation damage will result.

Although the mechanisms by which different types of ionizing radiation interact with the target material are quite different the net effect in all cases is the deposition of energy in the target usually in the form of kinetic energy of electrons and atoms. Often these collision products will have enough energy to cause further ionization or dislocations and by this avalanche process a single high energy particle entering a target can cause considerable biological, electrical or mechanical damage.

There are four major areas of concern which involve the radiation environment and its effect on space system and space crews:

- (1) Long term exposure to moderate electron and proton fluxes in the Earth's Radiation Belts. These effects include slow changes in properties of semiconductors, optical and other sensor surfaces, thermal radiative properties of materials. Another potentially serious problem for Space Systems is the accumulation of charge in insulating materials which can lead to internal and external electric discharges. The effects on biological systems particularly man can usually be controlled by fairly straight-forward shielding measures.
- (2) Solar Proton Events: As a result of major eruptions on the solar surface, intense, high energy pulses of protons lasting on the order of hours to days will reach the near-earth space environment. Both proton fluxes and energies can be several orders of magnitude larger than normal conditions with potentially severe consequences for both materials and personnel.
- (3) Galactic Cosmic Rays: Near-earth space is traversed by a very weak flux of extremely high energy nuclei originating outside the solar system. These give rise to Single Event Upsets (SEU) in which dense deposition of energy and subsequent separation of electric charge in a p-n junction can induce a change of state in an electronic circuit or memory element. Under certain conditions electronic components may be permanently disabled.
- (4) As a result of exo-atmospheric nuclear detonations a space craft may be exposed to intense pulses of x rays, gamma rays and neutrons. In addition there are important secondary effects such as artificial radiation belts and electromagnetic pulses which can have serious effects on space craft and their missions.

1 Radiation Thresholds

Ionizing radiation affects many components of space systems. The thresholds above which radiation effects must be considered vary greatly as shown in the table below:

Table 2. Radiation Damage Thresholds

Material	Damage Threshold (RAD)
Biological matter	10^1 - 10^2
Electronics	10^2 - 10^6
Lubricants, hydraulic fluid	10^5 - 10^7
Ceramics, glasses	10^6 - 10^8
Polymeric material	10^7 - 10^9
Structural metals	10^9 - 10^{11}

As might be expected biological matter is the most radiation sensitive but solid state electronics is not far behind. As electronic components become smaller and smaller this sensitivity is likely to increase. These thresholds do not mean that people or equipment cannot operate above them but their operation must be considered in equipment and mission design.

2 Radiation Dose in the Belts

In Figure 9.10 to 9.17 are shown a series of plots of the daily dose (due to electrons and protons) for satellite orbits for various orbital inclinations. All orbits are assumed to be circular. the multiple curves shown on each plot correspond to the absorbed dose in the presence of various thicknesses of shielding (aluminum) The altitudes cover the range from low earth orbit (few hundred km) to geosynchronous orbit (about 36,000 km)

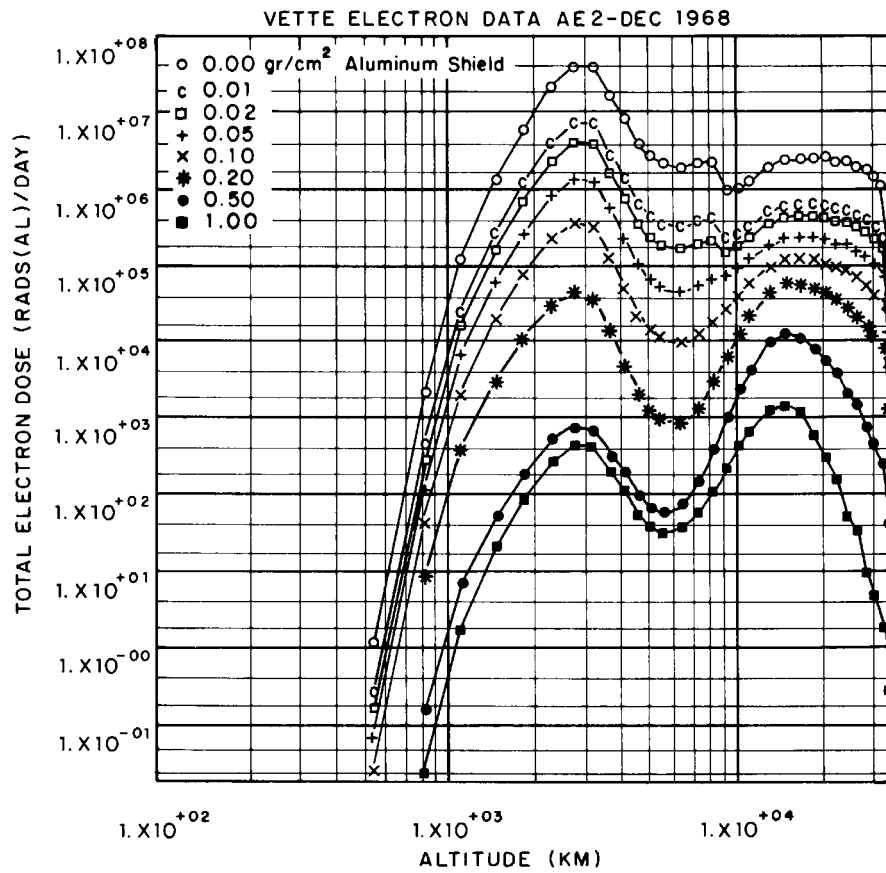


Figure 9.10 Daily dose (natural electrons) for a circular orbit satellite as a function of satellite altitude for 0° orbital inclination, with various shielding thickness.

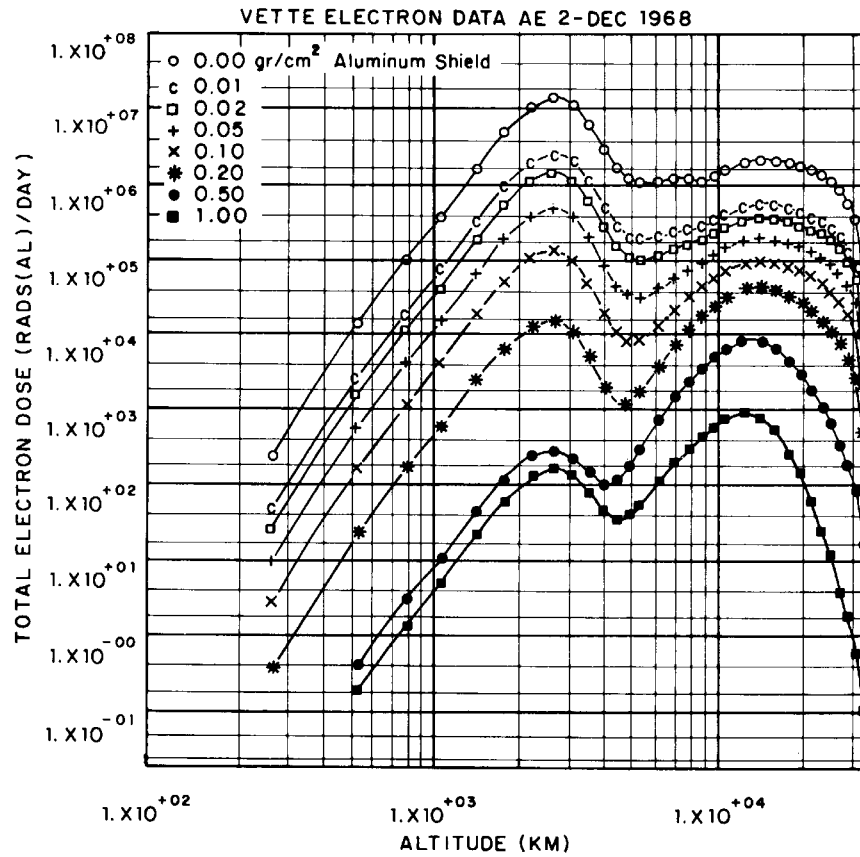


Figure 9.11 Daily dose (natural electrons) for a circular orbit satellite as a function of satellite altitude for 30° orbital inclination, with various shielding thickness.

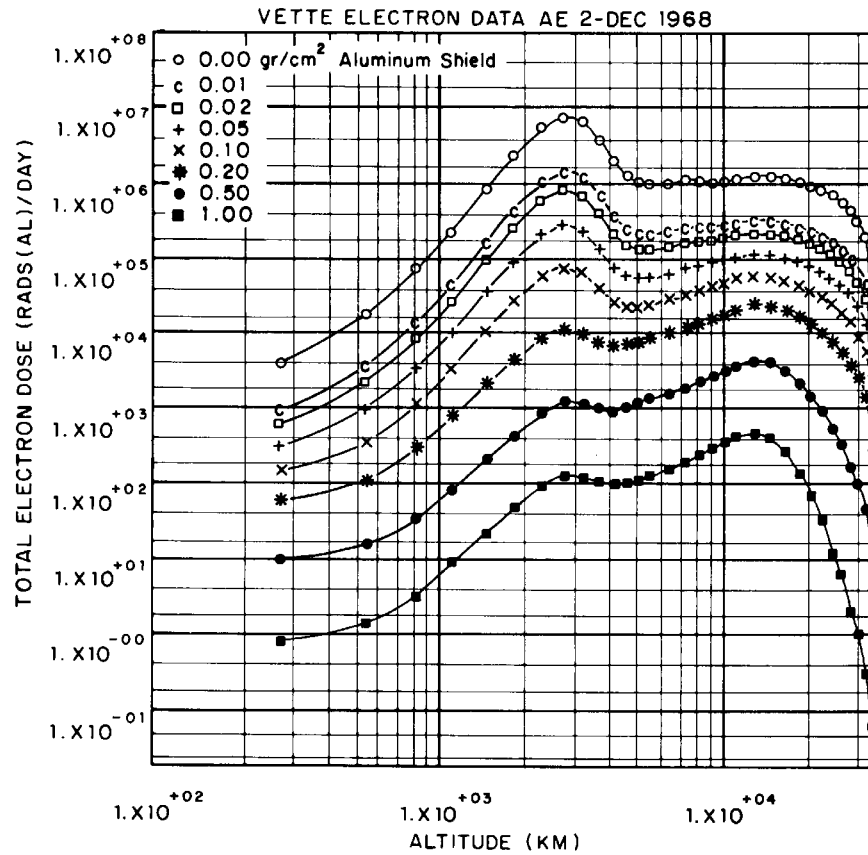


Figure 9.12 Daily dose (natural electrons) for a circular orbit satellite as a function of satellite altitude for 60° orbital inclination, with various shielding thickness

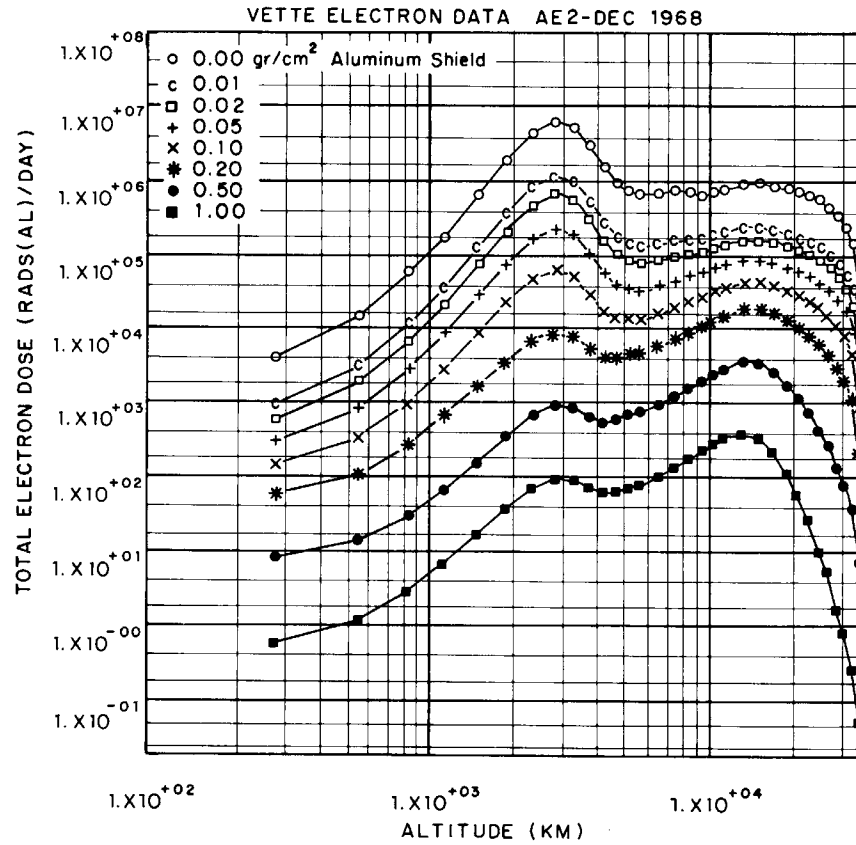


Figure 9.13 Daily dose (natural electrons) for a circular orbit satellite as a function of satellite altitude for 90° orbital inclination, with various shielding thickness

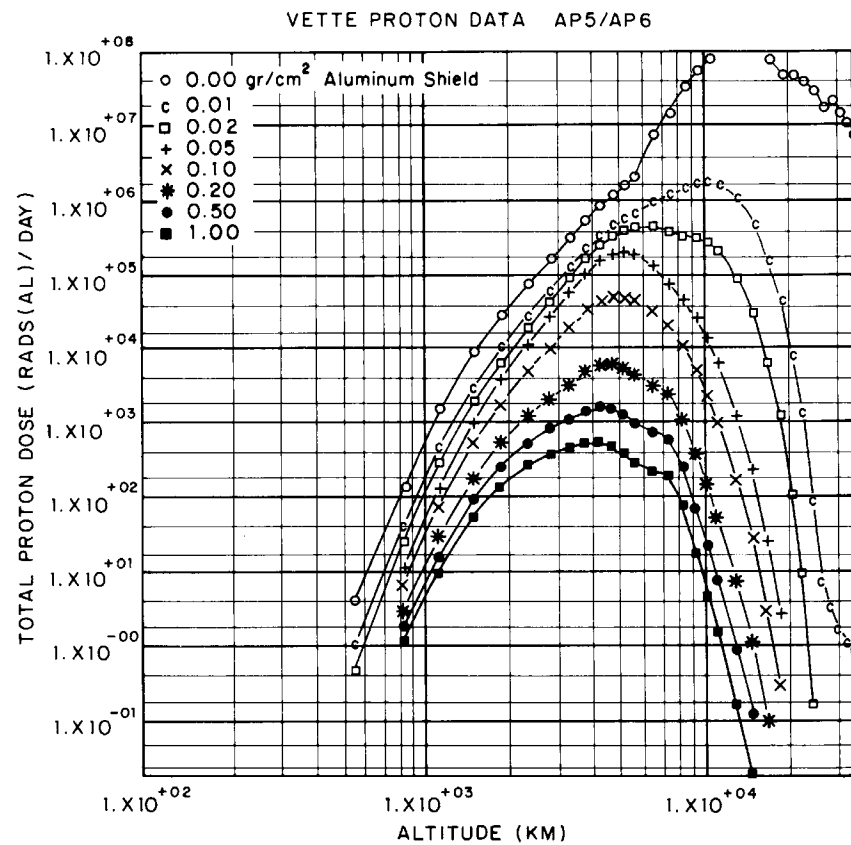


Figure 9.14 Daily dose (natural protons) for a circular orbit satellite as a function of satellite altitude for 0° orbital inclination, with various shielding thickness

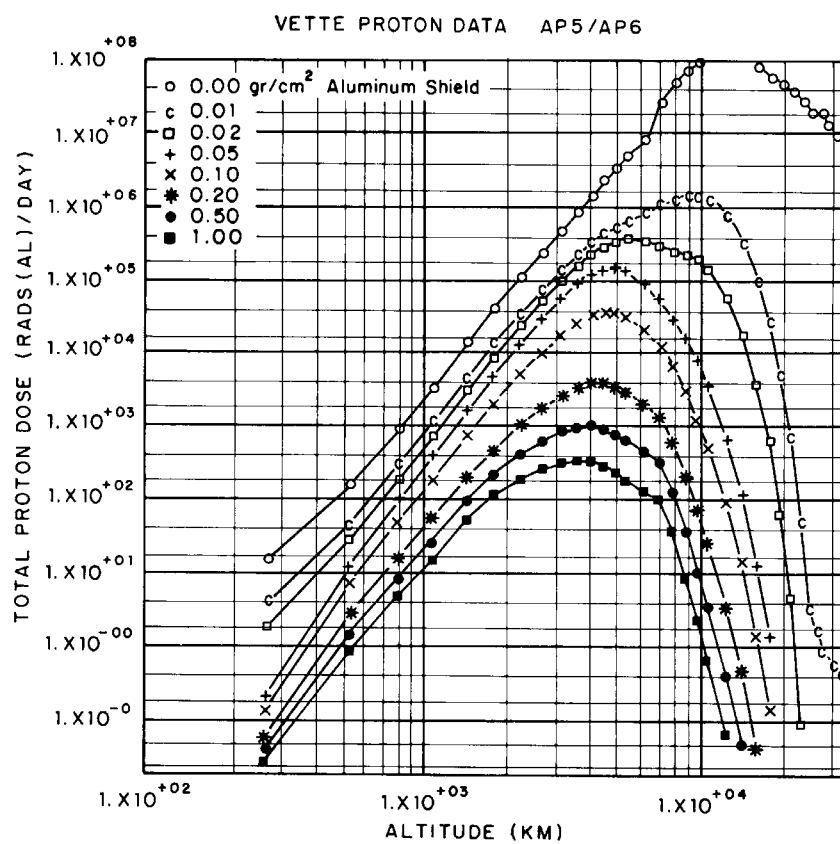


Figure 9.15 Daily dose (natural protons) for a circular orbit satellite as a function of satellite altitude for 30° orbital inclination, with various shielding thickness

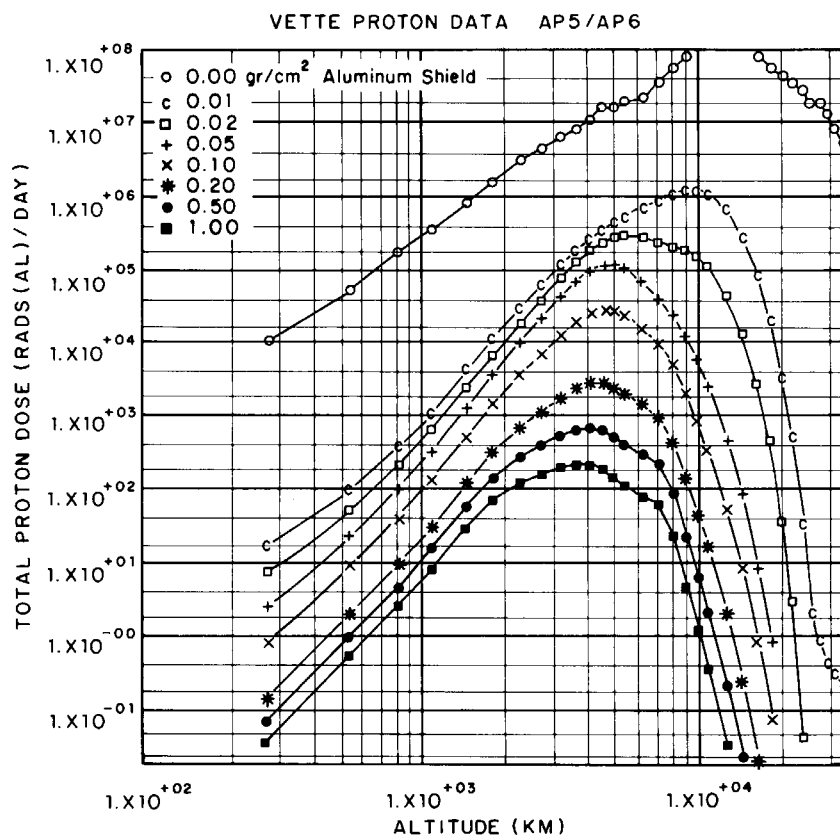


Figure 9.16 Daily dose (natural protons) for a circular orbit satellite as a function of satellite altitude for 60° orbital inclination, with various shielding thickness.

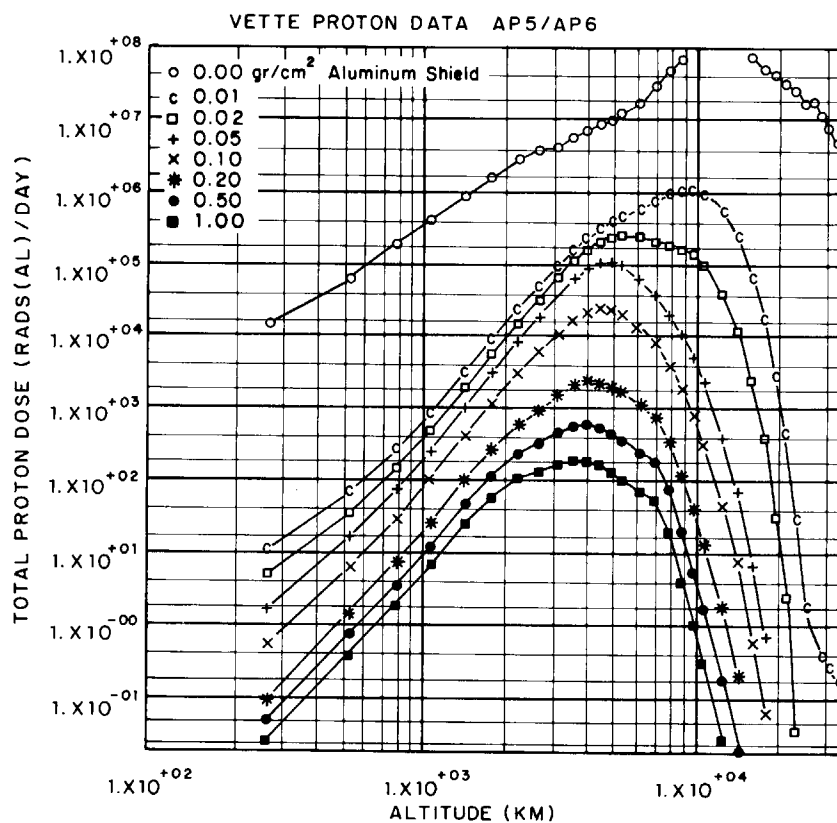


Figure 9.17 Daily dose (natural protons) for a circular orbit satellite as a function of satellite altitude for 90° orbital inclination, with various shielding thickness.

An example: Solar Cell Degradation

Solar cells are the most common power source for operational satellites, many of which operate in regions of appreciable ambient radiation. The power output of a solar cell may decline by 30%-40% over the lifetime of the satellite due to radiation-induced damage.

A typical solar cell consists of a n-p junction located about 0.2 microns below the surface with the n type material occupying the region between the surface and the junction and the bulk of the cell being p type material.

Exposing the cell to sunlight will generate free electrons mostly in the bulk p type material below the junction. Blue light is 99% absorbed within 0.2 microns of the surface where as red light penetrates about 200 microns before being 99% absorbed. Thus the majority of the free electrons produced within the cell must diffuse through the crystal to the n-p junction, be accelerated across the junction into the n type material and to the electrical contacts on the surface which connect to the external circuit. Typical solar cell efficiencies are about 11.5% meaning that 88.5% of the incident solar radiation is lost in the form of heat.

Radiation affects solar cells in two major ways:

- (1) Darkening of coverslides.
- (2) Decrease of charge carrier lifetimes.

(1) Coverslides: If a free electron is created in the coverslide it may be trapped by an impurity atom and form a charged defect called a color center. A color center is an atom-like structure with a set of energy levels and the trapped electrons can absorb solar photons passing through the initially transparent material. This process leads to a decrease in the incident solar flux and a reduced power output from the cell.

(2) Decreased carrier life time: Radiation can displace atoms in the crystal lattice forming positively charged interstitials. An electron diffusing through the crystal on its way to the n-p junction can get trapped into an orbit around an interstitial atom and can no longer contribute to the current flow. The average life time of the free charge carriers has been reduced by this process, meaning that fewer electrons will reach the junction and the output current will drop.

Figure 9.18 shows the effect of prolonged radiation on the power output of solar cell.

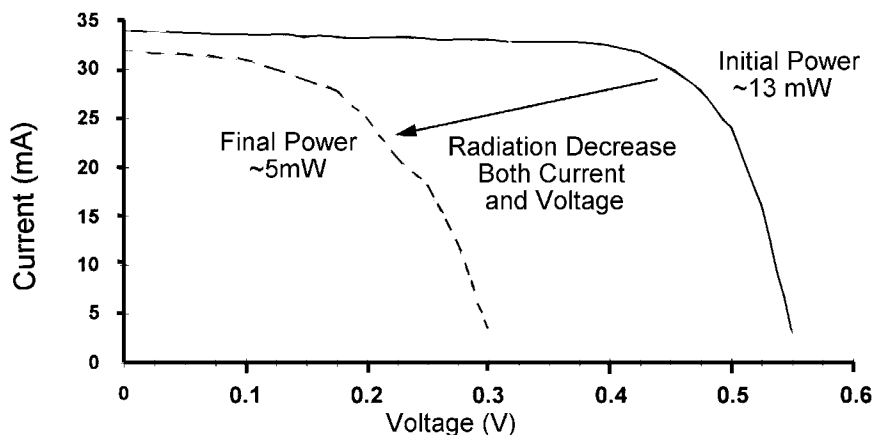


Figure 9.18. Effect of radiation dose on power production

To determine the effect of radiation on a particular solar cell we need to integrate over the energy spectrum of the radiation present. For ease of comparison it is customary to convert the actual fluences into equivalent fluences of either 1 MeV electrons or 10 MeV protons. That is, the damage produce by a distributed spectrum of particles is equated to the number of mono energetic particles that would be required to produce the same damage. The relative damage coefficients for electrons and protons are shown in Figure 9.19.

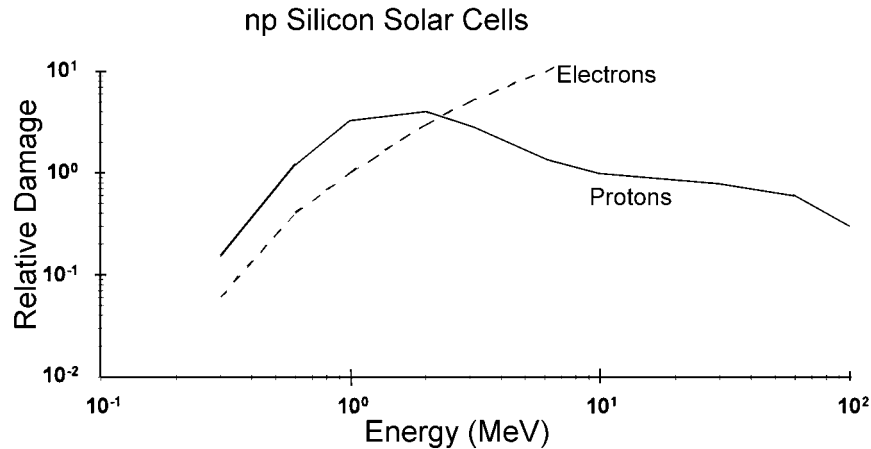


Figure 9.19. Relative damage coefficients for protons and electrons.

It should be noted that the absolute magnitude of the damage is quite different for protons and electrons: One 10 MeV proton produces the same damage as approximately three thousand 1 MeV electrons.

3 Solar Proton Events

Periodically the sun emits significant numbers of high energy protons (and some heavier ions) during severe solar disturbances usually referred to as Solar Proton Events (SPE). The length of the pulse is on the order of days with a range of hours to more than a week. Figure 9.20 shows typical doses near the earth due to a large SPE. There was a very large event in August 1972 and again in Oct 1989 and it is estimated that an astronaut on the moon shielded only by a space suit probably would have received a lethal dose of radiation.

To calculate the dose received from a fluence Φ we have that

$$\text{Dose (RAD)} = 1.6 \times 10^{-8} \left(\frac{dE}{dx} \right) (\Phi), \text{ where } \Phi = \int_0^t (\text{flux}) dt$$

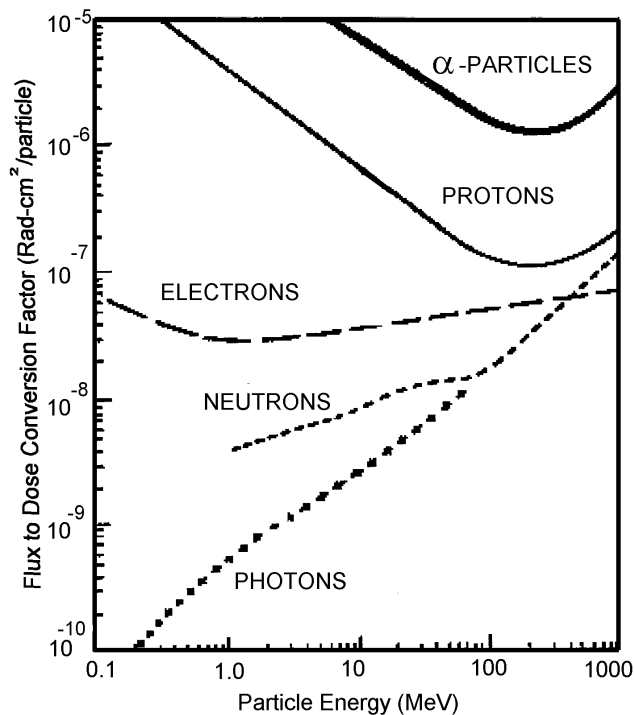


Figure 9.20 illustrates the fluence to dose conversion

Figure 9.21 shows the dose delivered by a range Solar Protons Event as a function of shielding thickness. Recall that 300 RADS is a 50% lethal dose

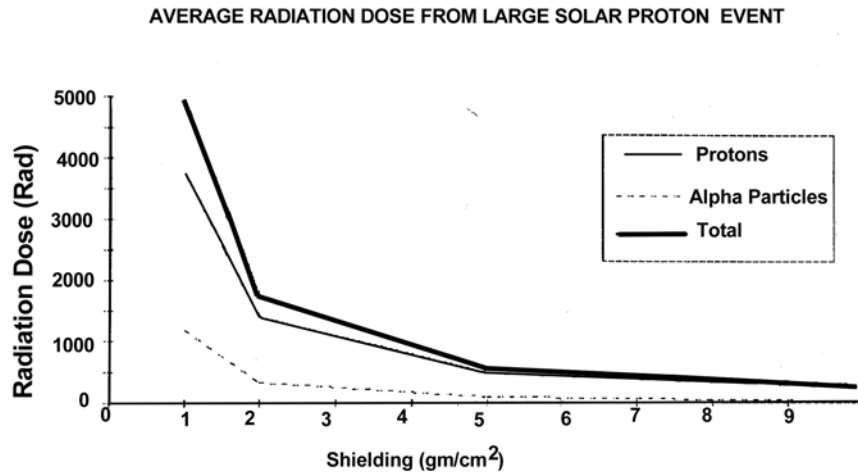


Figure 9.21

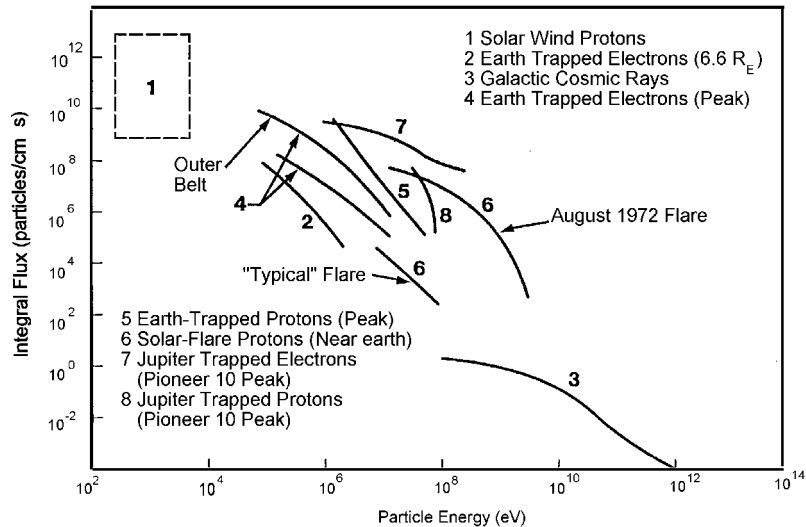


Figure 9.22. Radiation Environments of concern to spacecraft. This figure provides a summary of particle fluxes for trapped radiation as well as fluxes due to solar Proton Events and Galactic Cosmic Rays. JPL figure.

4 Galactic Cosmic Rays

There exists a low flux ($\sim 4 \frac{\text{particles}}{\text{cm}^2 \text{ sec}}$) of very high energy particles which appear to fill our galaxy isotropically. This flux is mostly protons (85%), He (14%) and 1% heavier ions with energies up to 10^{19} eV. The total dose due to these particles is quite small (a few RADs/year). They do present a problem by causing single event phenomena in electronics.

If a cosmic ray particle passes through or close to a p-n junction a current will be generated which is proportional to the dose rate

Many modern electronic devices have dimensions so small that the dose rate currents arising from the passage of a single energetic particle may be sufficient to alter the operating characteristics of the device in question. The resulting disruptions are called *single event phenomena* (SEP) or *single event effects* (SEE). An effect is classified as “soft” if the damage is transitory and the device can recover. An example of a soft error is the reversible flipping of a memory bit. An effect is classified as “hard” if the damage is permanent and the device is lost, such as an irreversible bit flip. Two specific examples of SEE warranting further discussion are *latchup* and *upset*.

Latchup is said to occur when the device is transformed to an anomalous state that no longer responds to input signals. Latchup is usually confined to bulk complementary metal oxide semiconductor (CMOS) devices. For typical integrated circuits the latchup threshold is on the order of 10^8 RAD/s. Upset occurs when a device is caused to function in a manner that is not consistent with its design characteristics. For example, the resulting localized electric fields and currents associated with radiation-induced currents may cause a memory register to change its state. The register is said to have been upset by the radiation. Upset thresholds are dependent on the design specifics of the device in question and may be as low as 10^7 RAD/s or as high as 10^{21} RAD/s. Upsets resulting from the passage of a single particle of radiation are termed *single event upsets* (SEUs). One characteristic of SEUs is that they are statistically guaranteed to occur in any device that proves to be susceptible to them. In the GEO environment upset rates may be as low as 10^{-10} errors/bit-day or as high as 10^{-4} errors/bit-day depending on the nature of the device in question. Upsets are directly dependent on the cross section, or LET, of the radiation for the device in question.

D High-Altitude Nuclear Burst Effects

An explosion of a nuclear weapon creates an environment which is extremely hazardous to life and equipment. In general, nuclear detonations can be classified into four broad categories: (1) air burst, (2) high altitude bursts, (3) surface bursts, and (4) sub-surface bursts. Although certain gross aspects of all four burst categories are similar, the high-altitude nuclear burst, sometimes referred to as an exo-atmospheric burst, presents the most significant danger to space operations. High-altitude nuclear bursts can seriously degrade communications systems, damage satellites, blanket large geographical areas with electronic and electrical equipment damaging radiation and electromagnetic energy, and expose humans to life threatening doses of radiation.

1 Characteristics Of A Nuclear Detonation

A nuclear detonation results from either the fission of isotopes of the elements uranium or plutonium, or, by the fusion of light elements such as deuterium. In both cases, the process releases large amounts of energy and radiation in a small amount of time and space. The altitude at which the blast occurs will, to a large degree, dictate how this energy and radiation is coupled into the surrounding environment. Figure 9.23 depicts the principal environments resulting from a nuclear explosion. In general, the principle effects of a nuclear detonation can be divided into blast and shock, thermal radiation, initial and residual nuclear radiation, and the electromagnetic pulse (EMP). Table 3 defines the effects of a nuclear explosion for different types of bursts.

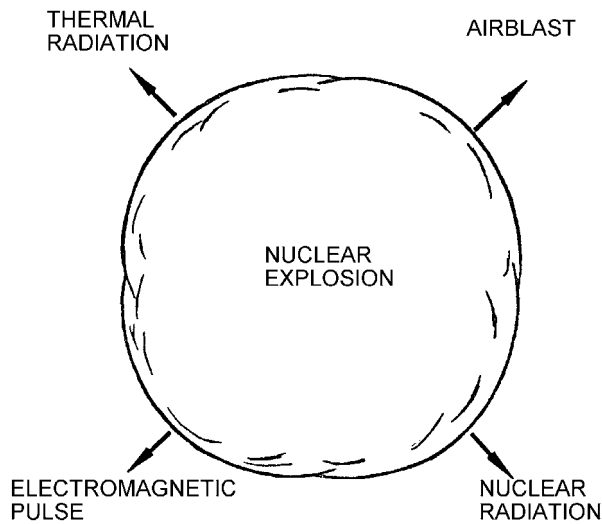


Figure 9.23 Principal Environments Resulting from a Nuclear Explosion


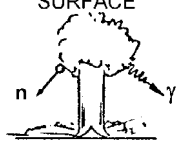

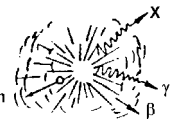
BURST ENVIRONMENT	UNDERGROUND	SURFACE	ATMOSPHERIC	EXO-ATMOSPHERIC
				
β -PARTICLES				✓
X-RAYS				✓
γ -RAYS		✓	✓	✓
NEUTRONS		✓	✓	✓
EMP		✓	✓	✓
BLAST	✓	✓	✓	
THERMAL	✓	✓	✓	
DUST/PEBBLE	✓	✓		
SHOCK	✓	✓		
CRATERING	✓	✓		
EJECTA	✓	✓		

Table 3. Summary of Burst Types and Environments

High-altitude burst are nuclear bursts which occur at altitudes above 30 km. In this region the scarcity of the atmosphere supports little coupling of the burst's energy into blast and thermal effects. The energy which is associated with a high-altitude nuclear burst is in the form of the initial prompt radiation, the kinetic energy associated with the debris, and the residual or delayed radiation. X-Rays account for between 75-80 percent of the initial radiation, gamma rays approximately .3 percent, and neutrons approximately 1 percent. The debris kinetic energy will account for between 15-20 percent of the total energy, and the residual radiation in the forms of gamma rays will account for approximately 5 percent of the total weapon output.

Unlike the lower altitude air, surface, and sub-surface burst whose effects are confined to a relatively small radius about the detonation point by the atmosphere, the effect associated with a high-altitude burst creates a post-detonation environment which is spread out over a wide area. The effects fall into four categories:

1. Ionization Effects in the Atmosphere
2. Electromagnetic Pulses (EMP)
3. Radiation Induced Damages in Space Systems
4. Artificial Radiation Belts

2 Ionization Effects In The Atmosphere

When a high-altitude nuclear burst occurs, the radiation will isotropically propagate outward from the detonation point. The volume of the atmosphere which is irradiated is known as the deposition region. Figure 9.24 depicts the formation of the deposition region following a high-altitude nuclear burst. The altitude to which the radiation will penetrate is known as the stopping altitude and is defined as the altitude in the vicinity of which a specified ionizing radiation coming from above deposits most of its energy by absorption in the atmosphere. For a large yield-weapon detonated just above the ionosphere, the deposition region will be about 1500 km in diameter and about 30 km thick. The prompt and residual radiations and their associated stopping altitudes are shown in Table 4.

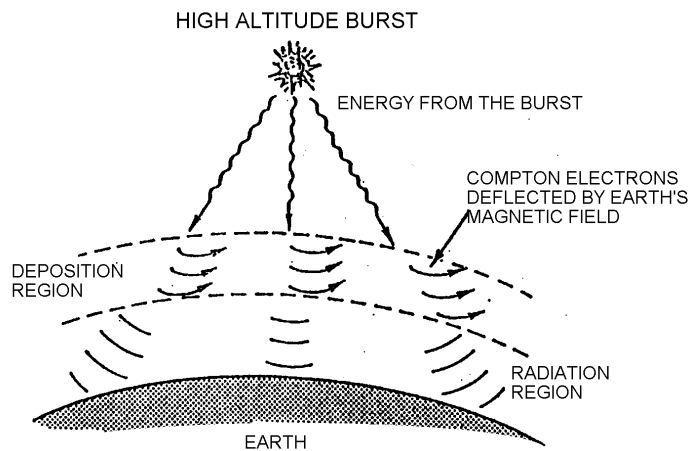


Figure 9.24: Formation of the Deposition Region

Table 4. Approximate Stopping Altitude for Radiation Causing Ionization

Weapon Output	Stopping Altitude (km)
Prompt radiation	
X-rays	60 to 90
Neutrons and gamma rays	25
Debris ions	110
Delayed radiation	
Gamma rays	25
Beta particles	60

The deposition region represents a large volume of the upper atmosphere which has been partially ionized by the prompt X-Rays and Gamma rays emitted by the detonation. A large number of free electrons are produced via the photoelectric effect and the Compton Effect.

The ionization of the upper regions of the atmosphere can seriously affect radar and communication systems. Electromagnetic radio and radar waves which pass through the deposition region will experience attenuation, signal distortions, and in some cases complete absorption. Satellite communication systems which rely on phase shifting or frequency shifting techniques may experience unwanted shifts in frequency and phase because of ionization induced changes in the propagation velocities of electromagnetic waves which pass through the deposition region.

3 Electromagnetic Pulse (EMP)

Nuclear explosions of all types are accompanied by the generation of an electromagnetic pulse (EMP) which is a sharp, pulse of electromagnetic radiation produced when an explosion occurs in an unsymmetrical environment. It is characterized by intense electric and magnetic field with rise times on the order of a few nanoseconds (10^{-9} sec), and decay times of a few tens of nanoseconds. The energy associated with EMP will be contained in this high intensity, short duration, electromagnetic wave with a sharp leading edge. This energy can be collected by a suitable collector and transferred to other components and equipment and since the energy is received in such a short period of time, intense currents are produced which can severely damage equipment.

There are two types of EMP which will be discussed, Deposition Region Generated EMP, and System Generated EMP.

4 Deposition Region Generated EMP

Under the proper circumstances a significant portion of the energy release during a nuclear detonation can appear as an electromagnetic pulse. The Deposition Region Generated EMP, usually referred to as just the EMP, is important for several reasons, listed below:

1. The effects of the EMP can have a lethality radius of thousands of miles.
2. The EMP is a fast rise time broad bandwidth pulse which poses a significant threat to almost all electrical and electronic equipment with the radius of lethality
3. Countermeasures to the EMP significantly increase the cost, weight, and complexity of space systems

The mechanism primarily responsible for the generation of the EMP is the Compton scattering of prompt gamma rays. As the gamma rays are scattered, the electrons which are emitted will have energies on the order of .5 MeV. These electrons will travel in a forward direction and will be influenced by the earth's geomagnetic field. They will travel along geomagnetic field lines in a spiral path until they are stopped by collisions with air molecules. The electron motion about the geomagnetic field lines produces a strong electrical current transverse to the propagation direction of the gamma rays. This transverse current radiates an electromagnetic wave which is the EMP. The fact that the forward scattered electrons travel close to the velocity of light causes radiation emitted at various altitudes to be approximately in phase thus greatly increasing the intensity of the final pulse. Table 5 gives a comparison of the electric fields generated by various sources.

Table 5. Electric Field Intensities (Ref. 3)

Source	Intensity (volts/meter)
EMP	100,000
nearby radar	200
nearby communications	10
Metro area	0.01

Unlike man-made (radar) and intense natural electromagnetic fields (lightning) which tend to be localized to a small area, the EMP is orders of magnitude higher and widely distributed. The EMP caused by a single high-altitude nuclear burst could cripple electrical and electronic systems for

thousands of miles from the detonation point. Figure 9-25 shows the effective radius of a typical high-altitude burst detonation over the United States.

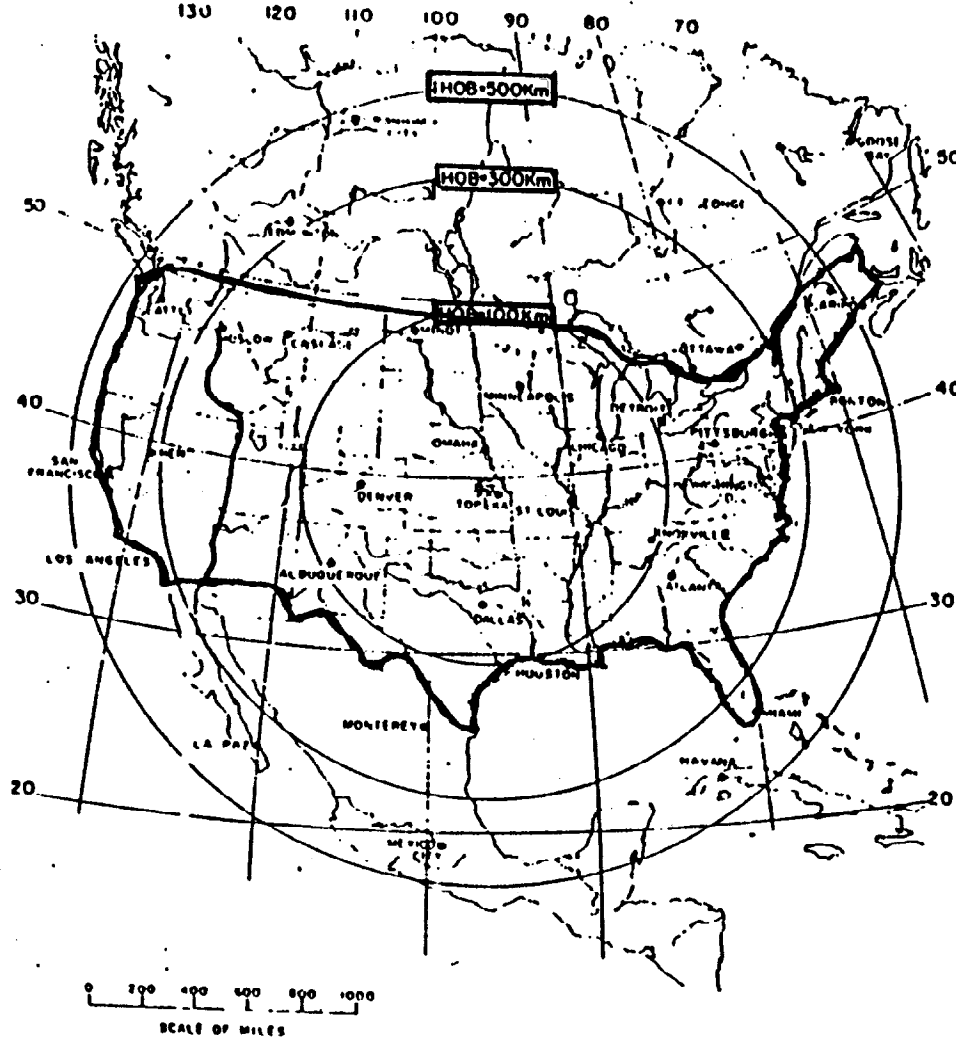


Figure 9.25. The Effective Radius of a High-Altitude Nuclear burst

5 System Generated EMP/Internal EMP

Another significant effect associated with a high-altitude nuclear bursts which is important to satellites is the electromagnetic pulse called the “system-generated EMP/Internal EMP” (SGEMP/IEMP). The mechanisms which are responsible for the creation of SGEMP/IEMP are similar to those responsible for the generation of the deposition-generated EMP. When the ionizing radiation from a high altitude nuclear burst, particularly gamma rays and X-Rays, interact with the body of a satellite, the irradiated material will release electrons due to Compton scattering and the photoelectric effect creating large electron emission currents and intense electromagnetic fields. This electromagnetic environment is coupled into the interior of the structure and is defined as SGEMP. In addition, since the satellite is basically a cavity, very high electronic fields will be induced across the interior of the spacecraft. These intense electric fields will in turn induce currents in the electrical and electronic components on-board the satellite. Figure 9.26 illustrates the generation of SGEMP/IEMP.

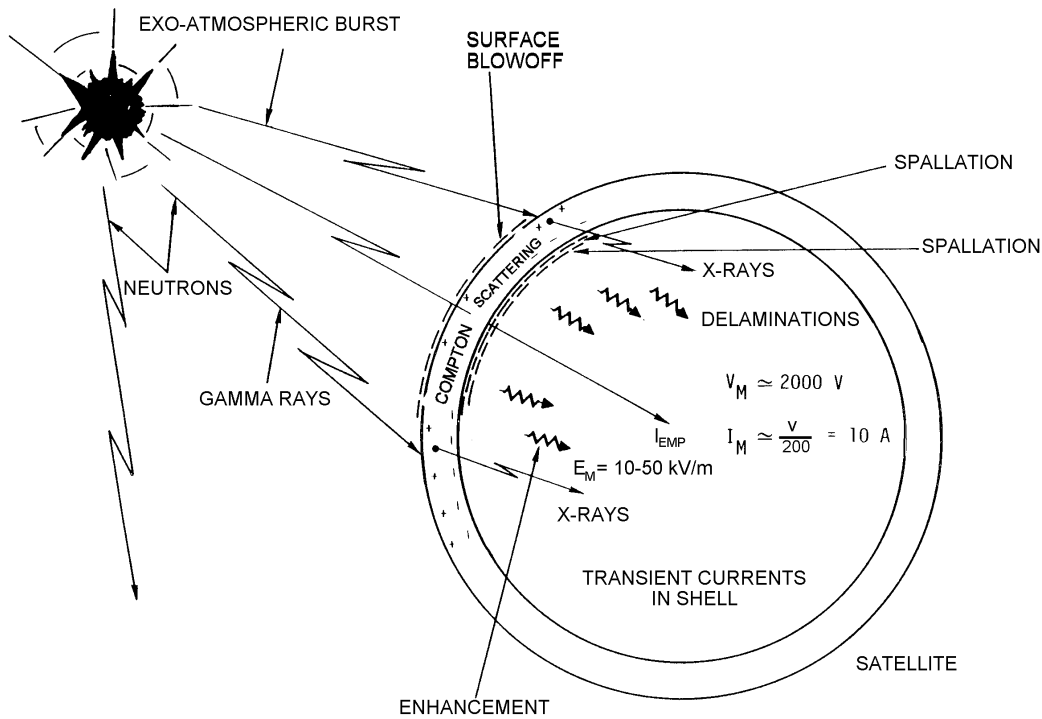


Figure 9.26. System Generated EMP and Internal EMP

6 Artificial Radiation Belts

Following a series of exo-atmospheric tests conducted by the U.S. and the Soviet Union in the 1960's substantially increased charged particle densities were observed in the Trapped Radiation Belts. Some of these artificial belts persisted for weeks to months and the population did not return to normal for about two years. A number of operational satellites were partially or totally incapacitated by the increased radiation encountered.

E Design Considerations Relative to Radiation

There are 4 categories of countermeasures

1. Shielding: Material surrounding sensitive components or placement of structures or less sensitive equipment to act as radiation absorbers
2. Part Selection: Utilize components which are known to be more radiation tolerant and have high resistance to latch up and upsets. Safety factors >5 are desirable if available
3. Redundancy: Oversize solar arrays and circuit redundancy including coincidence requirements and backup circuitry and parts.
4. Recovery Algorithms: Software capable of recovering the system from latchup or upsets

All of these measures add cost and weight but depending on circumstances they may be critical to the success of the mission

Blank page

Appendix 1 USEFUL EQUATIONS

EM Waves

$$\lambda f = c; \quad E = hf; \quad \lambda = \frac{hc}{\Delta E}; \quad c = 2.998 \times 10^8; \quad 1 \text{ eV} = 1.602 \times 10^{-19} \text{ Joules}$$

$$h = \text{Planck's Constant} = \begin{cases} 6.626 \times 10^{-34} \text{ Joule-seconds} \\ 4.136 \times 10^{-15} \text{ eV-seconds} \end{cases}$$

$$\Delta E(\text{eV}) = \frac{1.24 \times 10^{-6}}{\lambda(\text{m})} = \frac{1.24}{\lambda(\mu\text{m})}$$

Bohr Atom:

$$r_n(\text{meters}) = n^2 \times 0.528 \times 10^{-10} / Z.$$

$$E_n = -\frac{1}{2} \left(\frac{Z e^2}{4 \pi \epsilon_0 \hbar} \right)^2 \frac{m}{n^2} = Z^2 \frac{E_1}{n^2}; \quad E_1 = -\frac{me^4}{32 \pi^2 \epsilon_0^2 \hbar^2} = -13.58 \text{ eV};$$

$$number \propto e^{\frac{\text{Bandgap Energy}}{\text{Thermal Energy (kT)}}}$$

Black Body Radiation

$$c = 3 \times 10^8 \frac{\text{m}}{\text{s}}; \quad h = 6.626 \times 10^{-34} \text{ joule-s}; \quad k = 1.38 \times 10^{-23} \frac{\text{Joule}}{\text{Kelvin}}$$

$$Radiance = L = \frac{2 hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda kT}} - 1}$$

$$\text{Stefan Boltzmann Law:} \quad R = \sigma \epsilon T^4 \left(\text{Watts/m}^2 \right)$$

$$\epsilon = \text{Emissivity}; \quad \sigma = 5.67 \times 10^{-8} \left(\frac{\text{W}}{\text{m}^2 \text{ K}^4} \right); \quad T = \text{Temperature (K)}$$

$$\text{Wien's Law:} \quad \lambda_{\text{max}} = a/T \quad a = 2.898 \times 10^{-3} \text{ (m K)}$$

$$R_{\text{earth}} = 6.38 \times 10^6 \text{ m},$$